How to Invent a Language

by

David J. Peterson

Introduction

The first time I heard a language of mine spoken on screen was at a cast and crew premiere event for the first season of HBO's *Game of Thrones*. It was a lavish, but, comparatively speaking, poorly attended event. George R. R. Martin was there, but many of the seats reserved for cast members in the Ray Kurtzman theater remained vacant throughout the screening of the first two episodes of the series. Needless to say, they didn't know how big this thing would get (who did?), but I appreciated the extra leg room—and the front row seat.

My initial reaction to hearing Dothraki, the language of the long-braided, horse-riding warriors, though, was one of dismay. The first line one hears in the series is in the pilot, when Illyrio Mopatis, welcoming Khal Drogo and his band into his courtyard to arrange a marriage, says *Athchomar chomakaan*—"Welcome" when said to one person. I misremembered how I'd translated it, though, and thought he should have said *Athchomar chomakea*—"Welcome" when said to more than one person. So even though Roger Allem's performance was fine and it was I who was mistaken, I was a little miffed. After the screening had finished we all got in line to congratulate David Benioff and Dan Weiss on a successful premiere, and when they asked me how the Dothraki speakers did, my face must have betrayed me, for David said to me, "You know, we [i.e. he and Dan] were talking, and we realized, if any of the actors made a mistake, who would know it—except for you?"

This is actually a question I've gotten a lot since. That is if there's an actor performing in a created language that no one speaks, who will know if they make a mistake aside from the one who created the language? From experience, I can tell you

that the actors always know (and it frustrates them when the takes with errors make their way into the final cut), but let me focus on the audience.

If you, as a viewer, sit down and listen to one line from a created language and nothing else, it's nearly impossible to tell if it's a created language, a natural language (one that exists in our world), or gibberish—to say nothing about whether or not the actor gets all the words right. If that's the extent of the linguistic material in the production, it doesn't matter what work went into creating the line.

As the number of tokens increases, though, the odds of the casual audience member picking up on inconsistencies increases. It's not every fan that pays attention to what actors are saying in a language they don't understand, but there are those who do. Furthermore, television shows and movies aren't plays—that is, they aren't events that happen at one moment in time and are never seen again. If the general public is anything like me, *most* of the television and movie viewing they do now isn't done live—and if a show is worth its salt, they'll watch it again and again and again and again.

As a language creator, I always had a bit of a different perspective. When I was creating Dothraki, I wasn't creating it simply to fill out the requisite non-English dialogue. I had an idea that *Game of Thrones* could be big, and could occupy a special place in television history—just as George R. R. Martin's books already do occupy a special place in the history of fantasy. The work I was doing, then, would need to be something that would stand the test of time. Because even if a fan who's never heard of the books can't tell if one actor makes a mistake in the premiere on their first viewing, fans five, ten, twenty years from now *will* be able to tell. And, of course, if mistakes crop up, they won't belong to the show, the producers, or the actors: they'll belong to me.

*

When I was a kid, the original *Star Wars* trilogy had just completed its initial run in theaters, and *Star Wars* was *everywhere*. I had a toy sand skimmer (which I broke), a toy tie fighter (which I also broke), and a read-along *Return of the Jedi* picture book with accompanying record which would play the sound of a ship's blaster when you were supposed to turn the page. (If you're too young to be familiar with record players as anything other than "vinyl", type "PacMan record read along" into YouTube to familiarize yourself with the concept. That was my childhood.)

In short, aside from *He-Man*, *Star Wars* was pretty much *the* thing if you were a child of four in 1985. At that age, when I watched movies, I didn't really pay careful attention to the dialogue, and wasn't able to follow stories that well. Consequently when the *Star Wars* trilogy was rereleased in 1995, I rewatched it eagerly. Once I got to *Return of the Jedi*, I was struck by what I thought was a particularly bizarre scene. In the beginning, Princess Leia, disguised as a bounty hunter, infiltrates Jabba the Hutt's palace in order to rescue Han Solo. She pretends to have captured Chewbacca, and engages Jabba to negotiate a price for handing him over. In doing so, Leia pretends to speak (or evidently *does* speak, via some sort of voice modification device) a language Jabba doesn't. He employs the recently acquired C-3PO as an intermediary. As near as I can tell, this is how the exchange goes (transcription is my own; accent marks indicate where the main stress is):

LEIA: *Yaté. Yaté. Yotó*. (SUBTITLE: "I have come for the bounty on this wookiee.")

C-3PO relays this message and Jabba says he'll offer 25,000 for Chewie.

LEIA: Yotó. Yotó. (SUBTITLE: "50,000, no less.")

C-3PO relays this message and Jabba asks why he should pay so much.

LEIA: *Eí yóto*.

The above isn't subtitled, but Leia pulls out a bomb and activates it.

C-3PO: Because he's holding a thermal detonator!

Jabba is impressed by this and offers 35,000.

LEIA: *Yató cha.*

The above isn't subtitled, but Leia deactivates the bomb and puts it away.

C-3PO: He agrees.

Order is restored.

I want you to remember that I was in seventh or eighth grade at the time that I was rewatching this. I was not a "language" guy at that point by *any* stretch of the imagination. I never dreamed that a human could invent a language, and even if I had, I probably wouldn't have been able to come up with a good reason for one to do so. Furthermore, up to that point, I'd never studied a second language, and the prospect filled with me dread (I had enough trouble understanding my Spanish-speaking relatives who always spoke too fast for me).

But even so, I knew something was wrong here. How on Earth does Leia say the same thing twice and have it mean something different the second time? Even if we take C-3PO for an unreliable translator (he is quite loquacious, after all), that only applies to the last two phrases. How could one expect to have an unreliable *subtitle*? After all, subtitles are supposed to lie outside the world of the film. If you can't rely on a subtitle provided by the film's creators, how can you rely on anything?

In trying to resolve this conflict, it occurred to me that the only plausible explanation for this aberrant phenomenon is that the language itself was correct, but worked differently from all other human languages. In our languages (take English, for example), a word's meaning can be affected by the context it's in, but if you control for context, the word will always mean the same thing. Thus, if you're telling a story about your dog, and you use the word "dog" several times throughout the story, it will still refer to a fur-covered animal that barks and covets nothing so highly as table scraps. This is fairly standard and uncontroversial.

What would happen if a language didn't do that, though?

Take, for example, the word I have transcribed as *yotó* above. What if it changed its meaning over the course of a discourse? Naturally, one would have to define a discourse, but I think it's fair to consider this conversation featuring Leia, Jabba and C-3PO a single discourse, so we can leave that concern aside for the moment. What if the word *yotó* has several definitions? Specifically, what if the first time it's used in a conversation it means "this wookiee"; the second time it's used it means "50,000"; and the third time it's used it means "no less" (or the rough equivalent of those)? The same, then, applies for all other words in the language. That would resolve the ambiguity. How could one possibly use such a language? Well, they *are* all aliens (*Star Wars*, recall, takes place a long time ago in a galaxy far, far away). Maybe they're just better at this stuff than humans. Why not?

This was where my brain went while rewatching *Return of the Jedi* for the first time. At some future date I may have shared this with a friend, but if I did, the response

was likely an eyeroll. This quirk was just an unimportant detail in an otherwise fantastic movie. Why bother about it?

And so that's pretty much where my thought experiment died. I didn't take it any further, and no one was really interested, so I didn't think about it again until college.

But that, of course, was in a different time—a pre-internet time. Who does a teenager have to share news with other than their family, friends and teachers? Who do they come in contact with? In 1995, that's pretty much only the people who live near you and with whom you interact on a daily basis. How would you ever get a hold of anyone else? How would I have known that someone in the Bay Area, let's say—less than 500 miles away—had the same idea I'd had and also found that exchange interesting? In 1995, there was no way.

Then the internet happened.

Yes, the internet had been around for a while in 1995, but it wasn't a thing that just anyone could have access to. America Online changed all that. Pretty soon it became a thing to race home from school and go into a chatroom with a bunch of random people to talk about…nothing. And that was how we entertained ourselves—*for hours*. What a world, where you could chat with someone who lived in Lancaster, Pennsylvania about how Soundgarden rules!

As it turns out, though, I wasn't the only person to pick up on this. Another conlanger I'd later meet at the First Language Creation Conference, Matt Haupt, asked exactly the same question, and devoted a blog post to deconstructing that scene specifically. And *we* weren't the only ones. The Ubese language has its own entry on the Wookieepedia (yes, that's a thing) where contributors have written up an entire

backstory for the language that is, first of all, not a full language, and, ultimately, poorly constructed and not worthy of serious consideration.

So let me bring back David Benioff and Dan Weiss's question to me on the night of the *Game of Thrones* premiere. If the actors speaking Dothraki or High Valyrian or Castithan or whatever make a mistake, who would know but the creator? Who would care? The truth is probably one in a thousand people will notice, and of those that do, maybe a quarter will care. In the 1980s that amounts to nothing. In the new millennium, though, one quarter of 0.001% can constitute a significant minority on Twitter. Or on Tumblr. Or Facebook. Or Reddit. And on whatever other social media service is currently taking the internet by storm. To take a recent (at the time of writing) example, there was *Frozen* fanfiction and fan art circulating the internet *before the movie had even premiered*—and when it did premiere, it took a matter of hours for everyone to learn that Kristoff's boots weren't properly fastened, and that this was a big deal as it was disrespectful to the Sami people and their culture.

One of the most significant things about our new interconnected world is that the internet can amplify a minority voice exponentially. Yes, few people, comparatively speaking, will care if an actor makes a mistake with their conlang lines. But thanks to the internet, those few people will find each other, and when they do, they'll be capable of making a *big* noise. Every single aspect of every single production on the big and small screen is analyzed and reanalyzed the world over—and in real time. Every level of every production is being held to a higher standard, and audiences are growing savvier by the day. Language—created or otherwise—is no exception. In order to meet the

heightened expectations of audiences everywhere, we have to raise what we expect from languages created for any purpose. After all, if we don't, we'll hear about it.

*

Though it might seem like language creation is a recent phenomenon, with the success of shows like *Game of Thrones* and films like *Avatar*, the conscious construction of language is probably as old as language itself. The earliest record we have of a consciously constructed language is Hildegard von Bingen's *Lingua Ignota* (Latin for "unknown language"), which was developed some time in the 12th century CE. The abbess's creation wasn't a language proper, but rather a vocabulary list of just over a thousand words (most of them nouns). Hildegard developed this "language" for use in song, primarily dropping Lingua Ignota words into Latin sentences for, presumably, a specific kind of religio-aesthetic effect. The words, for the most part, look quite a bit different from either German or Latin, and feature an overrepresentation of the letter *z* (cf. *Aigonz* "God", *sunchzil* "shoemaker", *pasiz* "leprosy")—and she wasn't shy about creating words for concepts that were…less than holy (e.g. *amzglizia* "male pudendum", *fragizlanz* "female pudendum", *zirzer* "anus", *maluizia* "prostitute"—the full list is fascinating).

The inspiration for Hildegard's creation came, she believed, directly from God. The same is true for other projects found around the same time, such as Balaibalan, created in Turkey some time in the 15th century. The impetus for the creation of these languages was always external and supernatural. As far as we know, no one had yet created a language for any other purpose.

Around the 16th century, a new type of language began to emerge: the philosophical language. These languages were born of philosophers and scientists who saw problems with our languages (in particular, the arbitrary association between form and meaning), and sought to correct them. Of these types of languages, John Wilkins' philosophical language is likely the most famous example (though Ro from the 20th century is a significant improvement on the concept). Using an example reproduced by Borges, in Wilkins' language, if *de* is the word for an element, then *deb* is the first or primary amongst the elements (i.e. fire), and *deba* is a part of the first of all the elements (i.e. a flame). Cave Beck had a different take on a potential universal philosophical language in 1657 which made use of numbers. Taking his favorite example, *3*, which has to do with the concept of abatement, *p3* is a man who abates; *pf3* is a woman who abates; *r3* is abatement; *x3* is the act of abating, and so forth. Another favorite: if *q317* is bold, then *qq317* is bolder and *qqq317* is boldest, but there it stops. I think he really missed out here, as it would be incredible to describe a mighty warrior as *qqqqqqqqqqqqqqqqqqqqqqq317* (it just makes sense).

Generally, the goal for philosophical experiments such as this one was to perfect language for the purposes of science. If language can obscure intention, on account of metaphor, idioms and vagueness, then a precise language would be of vital and obvious value to the entire scientific community. As it happened, though, none of the philosophical languages from this era ever caught on, though it wasn't on account of a lack of effort on the part of their creators.

The entire character of the created languages discussion changed forever in the 19th century, though, with the advent of the concept of the international auxiliary

language (IAL). The philosophical languages were intended to be precise, but not necessarily easy to learn or use (indeed, using languages featuring categorization systems such as those employed by Beck and Wilkins proved quite cumbersome). An IAL, by contrast, is created to be as simple to learn and use as possible, so that it can be learned and used by large numbers of people across the world—those who would otherwise share no common language.

Though at the time of printing there have been easily over a thousand IALs produced, the two earliest successes were Volapük (1879) and Esperanto (1887). Volapük, created by German priest Johann Martin Schleyer, was the first IAL to gain major notoriety. Schleyer derived its vocabulary from English, German and French, and intended for the words to be both recognizable and easy to pronounce. He was marginally successful in this regard, with, for example, the name of the language itself coming from two English words: *vol* from "world" and *pük* from "speak". As the language gained a following, there were some who wanted to change the language (e.g. simplifying some of the tenses, getting rid of the vowels *ä*, *ü* and *ö*, adding *r*, etc.). Schleyer resisted any attempt to change the language, and insisted on maintaining complete control over its character and use. Splinter groups rose up creating knock-off versions of Volapük, and before long, the language had practically no following.

Part of its downfall, though, was no doubt due to the ascension of newcomer Esperanto, created by Ludwik Lejzer Zamenhof. In contrast to Schleyer's methods in presenting Volapük, Zamenhof published his initial grammar of Esperanto under a pseudonym (Doktoro Esperanto—whence the name of the language), and laid no claims on the language's use or distribution. He eschewed copyright, and said that the

language was his gift to the world. As a consequence, when splinter groups arose, it was the community itself, rather than Zamenhof, who decided what to do. This led to a major schism, with the majority of Esperantists adhering to Zamenhof's original conception of the language, and a minority splitting off and forming the Ido language: a modification of Esperanto. Both languages continue to enjoy success to this day, though Esperanto takes the lion's share.

The success of Esperanto, in particular, gave rise to innumerable attempts at a simpler or more international auxiliary language. General enthusiasm for the widespread use of an IAL was quelled primarily by the world wars, but interest in their construction has, if anything, increased over time, with new ones popping up online almost every month. None have achieved the success that Esperanto has, and it seems unlikely that any will ever do so—let alone achieve the goal of all IALs, which is to become the world's default auxiliary language.

The next wave of language creation began in the 20th century, and is known as the artistic language (or artlang) movement. Though early works of fantasy or satire would often feature bits of supposedly fictional languages (cf. Jonathan Swift, James Cabell, Lord Dunsany, E. R. Eddison), none of these snippets were languages in the proper sense—they enjoy no existence outside the books they're found in, and are largely haphazard or circumstantial in construction. The same is true of the work of Edgar Rice Burroughs, though in his works there's a bit more linguistic material throughout. The first widely-known author to use a more or less fully constructed language was J. R. R. Tolkien, who set the bar very high.

Unlike other authors before him—or most who would claim him as an inspiration afterwards—Tolkien was a language creator before he penned his major works. In a way, the languages themselves served as the progenitors to the tales. He understood that language itself is inseparable from the culture that produces it (or "mythology", as he puts it), and he felt that if the languages he was creating had no place to breathe, they wouldn't have any kind of vitality. Middle-earth became the place where his languages could live, and so his legendarium was born.

In addition to being the first person on record to create a full language for a fictional context, Tolkien also did something no other language creator had done to that point: He created a language *family*. Quenya and Sindarin, his two most famous languages, descend from a common ancestor, Quendian, and themselves have languages which have descended from them, and other languages to which they are related. This is precisely how natural languages evolve in our world, and would naturally be appropriate for a fictional setting that has an alternate history with any kind of time depth similar to ours on Earth. The concept, though, was a novel one, and not fully appreciated in Tolkien's time, as his works of fiction overshadowed his contributions to the history of conlanging.

There were other notable artistic conlang achievements in the 20th century (for example M. A. R. Barker's creations for the *Empire of the Petal Throne* role playing game, or Christian Vander's Kobaïan, used in song lyrics by his band Magma), but none attracted the attention that Tolkien's work did. It was only a matter of time before the phenomenon moved to film and television. Early films and shows would often use ad hoc invented vocabulary for fictional foreign nations (cf. *Danger Man*), or for languages

the film makers didn't want to bother to reproduce faithfully (cf. *Thoroughly Modern Millie*). A handful of films used actual Esperanto, with the most prominent (or infamous) example being the film *Incubus*, starring William Shatner. According to actual Esperanto speakers who've seen it, the Esperanto is atrocious. Otherwise there was no noteworthy conlang work in a film or show before 1974's *Land of the Lost*, which featured an invented language called Paku (often erroneously referred to as Pakuni). Paku was created by UCLA linguist Victoria Fromkin, who was hired specifically for this purpose. Though a few dedicated fans have attempted to take down what little they can hear of the language in the show, the producers clearly didn't think it was in their best interest to publicize the language the way conlangs attached to modern franchises are publicized, which is a pity. The creation of the Paku language was a momentous event in the history of conlanging, for unless there are other instances which haven't yet come to light, this was the first time in history that an individual had been hired to create a language.

The idea of hiring out for a language was repeated a few times in the late 20th century—most notably with Klingon, which was fleshed out by Marc Okrand for *Star Trek III: The Search for Spock*, but other notable examples exist, like Matt Pearson's Thhtmaa language for NBC's *Dark Skies*—but the wave really started to grow with the advent of the internet.

The first gathering of language creators—either virtual or in person—occurred on July 29th, 1991. That's the date of the first ever message sent to the Conlang Listserv: an online listserv dedicated to those who created languages. The original members of the listserv met on Usenet, and decided to create their own listserv just for language

inventors. When they created it, the listserv needed a name, so they took the first syllable of "constructed" and the first syllable of "language" and created "conlang". Thus the word "conlang" was born. At the time, there were a number of competing terms for an invented language—planned language, model language, artificial language, created language, ideal language—but conlang was the one that eventually won out. The portmanteau strategy proved useful for coining words for other types of conlangs, and so new terms began to spring up: artlang, loglang, engelang, auxlang, altlang, lostlang, jokelang… At this point, -lang could probably be considered a fairly productive derivational suffix.

The constructed languages listserv was originally run from the Boston University physics department by a fellow named John Ross, before it was moved to the Datalogisk Institut in Denmark, where it stayed for a bit. When it could no longer be housed there, though, David Durand, one of the original members, used his alumni connections from Brown University to get the list a permanent home there, where it remains to this day.

Though the original list members probably didn't realize it at the time, the founding of the Conlang Listserv (hereafter Conlang) was a momentous occurrence in the history of language creation. Aside from contentious gatherings devoted to how to improve a specific auxiliary language, there had never before in history been a place where those who created languages would discuss strategies for doing so. For the first time language creators could compare their work to something other than Tolkien's languages or Esperanto and its many imitators.

Of course, like any community, Conlang had its rocky moments. While many Conlang members were interested primarily in sharing their own work and learning from others who would share theirs, a sizable contingent were advocates of one of a number of IALs: Esperanto, Novial, Volapük, or creations of their own. Arguments would often break out over which was the best language, and which should be supported as the one language that all the world should speak. As a result of the constant bickering, in 1996 a separate listserv was created—the Auxlang Listserv—and advocacy of *any* language was banned from Conlang. Instead, discussion on Conlang would be specifically devoted to sharing non-auxiliary language work, and discussing strategies related specifically to language creation.

This proved to be a real turning point for the craft, as Conlang began to accrue more and more members and develop its own traditions hereafter. In the summer of 1999, Irina Rempt, creator of Valdyan, initiated the first conlang relay. In a conlang relay, the first participant writes a text in their conlang, and then passes on the text, grammatical notes and a lexicon to the next participant. This next participant has forty-eight hours to use those materials to decode the text and then must translate it themselves into their own conlang, passing the text on with translation materials, and so forth, until all participants have had their turn with the text (called the "torch"). The resulting texts are often ridiculous, which is part of the fun. Here, for example, is the original text of the thirteenth conlang relay, which was composed in Da Mätz se Basa by Henrik Theiling:

De Mätz se Basa Line: An ein Muin äna Monat Wöpf kan ana da Drot s ano Bant fona Bos bänti s Urt da lei se Zän is sä ze.

One morning in December I could watch the following scene across the street from a bus stop:

A boy, obviously on his way to school, stood in front of a garden wall. He had an enormous red tomcat on his shoulder and this animal balanced skilfully while the boy tried hard to shake him off. With increasing panic he watched the bus stop knowing the bus would be due any second now... The cat was quite undisturbed. Once, he put a paw on the wall, but seemingly it felt too cold, so he quickly retreated to the boy's backpack. The moment when he was comfortably sitting, having tucked up his legs, the school bus arrived.

I suppose the cat spent quite an interesting day.

Now here's the final translation into Minza by Herman Miller:

Minza Line: Vyø jenzelu kaikat, seła køvu nintel vonyli rukumen røġisit.

Before the sun came out, a messenger woman was walking along bringing messages in a cart. She paused briefly as the cart arrived near the path which led to the market place, but then she disturbed an animal, which growled at her. The messenger woman held up a branch of a tree, and approached near the cougar.

Wanting to avoid danger, the messenger woman was anxious about the cougar, so she stopped far from him, but tried to obtain his trust. She gave a portion of food to the playful cougar. She thought he would approach the cart, but the shameful cougar wanted more food! It began to

rain, the woman holding the branch, so she wanted to stay far away from the cougar.

With the branch of a tree, she protected herself quite well from the cougar. From there the messenger woman reached the end of the path with her cart. After that day when she set out, the growl of the cougar has made more women and more men nervous.

The stories are absurd, but the practice is useful. Members of Conlang have found dozens of ways to test out their languages over the years, improving them as they go along. For example, translating the "Babel Text" (Genesis 11:1-9) became a standard stress test for a new conlang, thanks to the website Langmaker.com, created by Jeffrey Henning, which hosted various conlangs' translations of the Babel Text.

Conlang continued to grow and expand, until soon splinter communities began to emerge—including communities which held discussions in languages other than English (e.g. Ideolengua, the first Spanish-language conlang community). The various communities attracted some of the best language creators on the planet, and also produced a new generation of conlangers (of which I was one) whose conlanging heroes weren't Tolkien or Zamenhof or Okrand, but other conlangers like Sally Caves, Sylvia Sotomayor and Matt Pearson—names widely known and respected within the conlanging community, but virtually unknown without. To date, the best languages ever created were not created for television series or movies, but were created just for the joy of it—languages like Sally Caves Teonaht, Doug Ball's Skerre, Sylvia Sotomayor's Kēlen, Matt Pearson Okuna (formerly Tokana), Andrew Smith's Brithenig, John

Quijada's Ithkuil, Carsten Becker's Ayeri, and David Bell's ámman îar, to name a few among hundreds.

A key feature of some of the best languages I've seen in that time—including all those listed here—is they changed crucially as a result of contact with the community. This is the natural result of community in the most abstract sense. Consider: What artist never looks at any paintings but their own? What musician never listens to any music but the music they create? Yet this was precisely the state of conlanging prior to 1991. Some will have heard of Esperanto or Tolkien or Klingon, but a majority believed that they were the first person *ever* to create a language. For example, I believed I was the first one ever to create a language for purposes other than international communication —and that was in 2000. Part of this was due to a general lack of awareness, but part is also due to the nature of the activity.

Hobbies are what they are. People do things because they find them fun. Some hobbies can eventually lead to something bigger (painting, writing, sculpting, sports, etc.). Conlanging is a hobby that, even as late as 2009, no one in the conlanging community believed would amount to anything real. Part of this comes from older conlangers' personal histories. Parents who found their children creating languages would consider the practice so bizarre that they believed it to be indicative of some sort of mental disorder. Esperantists and other auxlang advocates considered the practice counterproductive and silly. Linguists would, at turns, either dismiss or deride the practice (one of the earliest works by a linguist which discusses conlanging in any depth is Marina Yaguello's 1984 work *Les Fous du Langage*, or *Lunatic Lovers of Language*, which is precisely as insightful as it sounds). On Conlang, members took to talking

about their language creation using the same terms homosexuals would to talk about their homosexuality. Revealing to one's parents, friends or colleagues that one conlangs is still referred to as "coming out".

The truth, of course, is that language creation is just a thing (and I mean that in *the* most prosaic sense). Conlanging is an activity that harms neither the conlanger nor the world around them. If anything, there's an intrinsic benefit in engaging with linguistic material—created or otherwise—as it exercises the parts of the brain that use language, and gives the conlanger a bit more of an open mind when they encounter languages other than those they speak natively. Even so, a lifetime of negative feedback has left its mark on the community, which has been tolerant of praise, but allergic to criticism, constructive or otherwise.

This lack of judgment was a key feature of the early Conlang community. Reactions to a piece of art are subjective, of course, but just as with any activity that requires any amount of skill or ingenuity, certain elements of language creation can be measured objectively. The community has routinely rejected any calls for objective measurement of any kind—or any criticism other than positive feedback. The thinking was that if this was a place where a conlanger could actually be themself, it should be a zone free of judgment (and certainly memories of the auxlang flamewars [i.e. arguments about the respective merits of the various IALs] are still fresh in the collective memory of the community). In many ways, this has helped to foster growth, but it's also hindered the development of the craft.

What I would advocate is a goal-driven approach to conlang evaluation. That is, when one creates a language, one creates it for a number of reasons. Those reasons

will determine the character of the language and the nature of what would amount to acceptable criticism. For example, if one creates a language for personal use and for no other reason, then the only acceptable criticism will come from the creator themself; all other criticism is immaterial. That shouldn't be true of a language created for television or film. The constraints will differ, as paid language creators have to bow to the whims of producers, directors and writers, and generally don't have as much time to develop their work as a conlanger working for themself, but within those constraints, criticism is appropriate—and healthy for the community. If a language has been developed for a fictional race of people in an otherwise realistic setting, the language should likewise be as realistic as possible. Does such a language appear to be naturalistic in all respects? If not, how could it be improved—how it might it appear more authentic? These are good questions to ask, and good examples to learn from.

Which brings us to today—and the conlangers of today. Up until, say, around 2004, I could confidently say that if there was anyone that had even dabbled in language creation, I had heard of them and of their language, and could list a couple of key traits of that language. The community was tight-knit, and even though it had branched off a bit, everyone still was able to keep tabs on everyone else, for the most part. In 2015, this is beyond impossible. Not only is it impossible to know every language creator, it's not even possible to know every language creation community. There are conlangers active on Tumblr, Twitter and Facebook that have no connection to any of the original communities—and they're drawing inspiration from languages that didn't exist even five years ago. And while this is great, the newest conlangers lack any means of evaluation or instruction. They know the word "conlang", but have never heard

of the Conlang Listserv. They know Na'vi, but have never heard of Moten. They've never had to defend their work as not being a serious attempt to create a new universal language (the first accusation most artlang creators faced when presenting their new language on the web). They're a new breed, and share none of the same assumptions that the early Conlang crowd did. This means they don't have the hangups older conlangers do, but also that they lack the history—and, most importantly, the knowledge—accrued over years of steady interaction.

When I get an e-mail from someone who's eager to create their own language and wants to know where to begin, I have a tough time explaining what it is they should do. I think my conlang education was good, for what it was, but what did I do? I joined the Conlang Listserv and spent a decade there sharing my work, learning from others, and learning more about language. Should everyone have to do the same? Where is the collected wisdom of the early conlang community? Why is it not written down somewhere that if you're creating a naturalistic ergative language, it will most likely be split ergative, and that those splits will happen in one of a small number of likely places in the grammar? This is something that every conlanger knows or eventually learns, but the information is only passed via word of mouth—it's like we're living in the 1300s, but we also have the internet and indoor plumbing!

This work is a sincere attempt to give new conlangers a place to start. It won't answer every question (no single book could hope to hold every scrap of information one needs to know to create a good language), but it should allow new conlangers to get a sense of the craft, and avoid having to reinvent every wheel that the conlang community as a whole has created and perfected over the last quarter century. My aim

is to help conlangers avoid expending mental energy on some of the nuts and bolts of language creation so they can focus on the more important question: What do I want to say with this new language that I can't say in my native language—or any other language that currently exists?

*

Before I get into the meat of the text, let me discuss some top level terminology that will apply to the work as a whole. Many battles have been won and lost on the internet when it comes to conlang terminology, but we've reached a point where there is broad agreement about certain terms, so I'd like to set them down here in print in order to add some stability to the debate (Oxford English Dictionary, pay attention). Here are some terms you'll need to know going through every section of the book (other terminology relevant to specific chapters will be discussed in those chapters):

- Conlang: A conlang is short for "constructed language", and is the consensus term for a created language. It was coined some time before the Conlang Listserv was founded in 1991, and has increased in use as the main term for a created language since then. Any language which has been *consciously* created by one or more individuals in its fullest form is a conlang, so long as either the intent or the result of the creation process is fully functional linguistic system. This includes Esperanto, Quenya, Dothraki, Lojban and Lingua Ignota, but doesn't include modern revitalization projects like Modern Hawaiian, Modern Cornish and Modern Hebrew—nor does it include creole languages like Tok Pisin, Bislama or Saramaccan.

- Natlang: A natlang is a natural language (both terms are in regular use), which is any of the languages that happen to exist in the world and evolved naturally. This includes any spoken language, creole or otherwise (Spanish, Ainu, Moro, Estonian, Kituba…), as well as signed languages (ASL, FSL, BSL, TSL, etc.). It also includes revitalization projects like Modern Hawaiian and Modern Hebrew, and dead languages like Latin and Akkadian. The point of emphasis is the nature of the origin of the system. Languages that began their existence by an act of conscious creation will share important features in common with other created languages that they won't share with natural languages. Consequently, even though Esperanto is now spoken natively by speakers all over the world, it's still important to understand that it began as a created language, and that, as a result, there's a reason it looks the way it does.
- Fictional Language: A fictional language is a language that's supposed to exist in a given fictional context, but which doesn't actually exist. For example, in the *Star Wars* scene I referred to above, Leia is supposed to be speaking the Ubese language. In our world, though, this language doesn't exist (its "construction" being entirely haphazard and never fleshed out). There's nothing to be said about the actual system, but in the fictional context, it's supposed to be understood to be an actual language.
- Real Language: A real language is one that actually exists, regardless of its status. Conlangs and natural languages are both real languages, since they actually exist in our world—or exist to the fullest extent that a language *can* exist. (After all, languages aren't objects: they're ideas, or patterns of behavior.

If all English speakers remain silent for a few minutes the world over, the language still exists.)

- Fake Language: Like Ubese, a fake language is one that's meant to give the impression of a real language in some context without actually being a real language. For example, I could tell you that I went to Finland and saw a sign that read *Kioriluvinen tääriällinö hänäskä*, which means "Leave your skis outside", and unless you knew I've never been to Finland, you might think, "Yeah, that's Finnish, I guess." Of course, it's entirely made up. It's based on nothing but my knowledge of the phonology of Finnish. Consequently, it's fake language—or fake language material. A conlang is *not* a fake language. It may not be large, but that just means it's not finished yet. Do not call a conlang a fake language. Those who do only make themselves look foolish.
- Code: I'll also include here "cipher" and "language game". All of these are systems that are created to hide meaning from those who don't know the system—like the message about drinking Ovaltine in *A Christmas Story* (oh. Spoiler alert). A key feature of all codes and ciphers is that they must *crucially* depend on another language in order to work. Take Pig Latin (e.g. ake-tay ig-pay atin-lay). No one will "get" Pig Latin if it's based on a system no one knows (e.g. iss-ray azdozgit-may ifthuliar-kay). The point is to have an in group and out group, with the in group understanding the trick that will help them turn the cipher back into the original language. Consequently, none of these things are conlangs. Pig Latin, for example, is nothing more than a bizarre way of speaking English.

- Jargon: Once every couple a years a news report will emerge about a company where they speak their own "language". What all of these reports end up pointing out is that any community of individuals will come up with a subset of vocabulary particular to their environment and experiences, and outsiders won't know this vocabulary. This is why those who've seen *Office Space* will know the term "TPS report", and those who haven't won't—or how a Hester is a term for returning a successful roll (or "through") in Gentlemen's Roll, a game my friends and I invented (patent pending). Crucially, no matter how many terms are invented, the language that they're used in is English. I can say something like, "We thoroughfared seven bolsters before the LFC structural last middleseven", and everyone can see that it's still English. No one will know what an LFC structural is, but it's clearly some sort of event. You have to be able to speak English in order for any of this to work, though, so vocabulary subsets like these are not actually separate languages.
- Dialect: An instantiation of a language is a dialect. Every human on Earth speaks a dialect of their own language. Your particular version of your language is called your idiolect (for example, I pronounce "both" with an "l" [i.e. bolth], and always have. Why? Because I do. Deal with it). A dialect isn't a "type" of language—i.e. it's not as if some people will speak the English language, and others will speak a dialect of English. *Everyone* who speaks English speaks a dialect of English. That some dialects are more highly prized than others is an unfortunate byproduct of social inequality and history. In the eyes of linguists,

all dialects are equal, in that they all achieve the functional requirements of linguistic interaction.

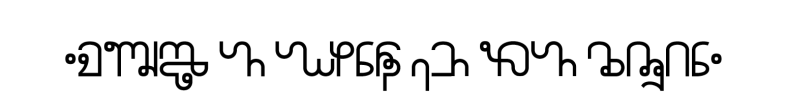
- Artlang: Short for "artistic language", this is a conlang created for aesthetic, fictional or otherwise artistic purposes.
- Auxlang: Short for "auxiliary language", this is a conlang created for international communication (or sometimes for communication amongst a specific subset of the populace, as with Jan van Steenbergen's pan-Slavic auxlang Slovianski).
- Engelang: Short for "engineered language", this is a conlang created to achieve some specific type of linguistic effect (e.g. to create a language without verb, as with Sylvia Sotomayor's Kēlen, or to create a language that uses LIFO grammar, as with Jeffrey Henning's Fith).
- Conlanger: Someone who creates a language—especially someone who engages in language creation regularly.
- Linguist: An individual employed by a linguistics or philology department who engages in the scientific study of language. (Important: This is *not* a synonym for "conlanger".)
- A Priori: An a priori conlang is one whose grammar and vocabulary are not based on existing languages. Dothraki is an example of an a priori conlang.
- A Posteriori: An a posteriori conlang is one whose grammar and vocabulary are drawn from an existing source. Esperanto is an example of an a posteriori conlang.

- Translation: Translation is the practice of rendering the content of a clause in one language into a different one. Here, for example, is a sentence of Castithan translated into English:

The boy appreciated the flower.

Translation does not take into account anything about the original language: It simply renders the meaning in a different one.

- Transcription: Transcription takes the text of one language and puts it into a form that's readable by a person who speaks a language that uses a different orthography—or into a neutral orthography. For example, here's the same Castithan sentence from above first transcribed in a Romanization system that can be more or less understood by English speakers, and then in the International Phonetic Alphabet, used by linguists and dictionaries throughout the world:



Fahazwa re rutsaye'ke zwore giopsa.

ˈfa.ha.zwa.ɾe ˈru.tsa.je.ke ˈzwo.ɾe ˈgi.o.psa

Transcription may take into account some of the conventions of the original language's writing system or its phonology, but it is *not* a translation. It's simply a method of rendering the same text in a different writing system.

- Gloss: A gloss (also referred to as an interlinear) gives the reader an idea what each word in a clause means, and/or what role it plays in the sentence.

Combining actual language data with a transcription, translation and gloss is crucial in determining not only what data in another language means, but *how* it means what it means. Below are two different ways of glossing the same sentence of Castithan used above (the latter with more detailed information than the former):

Boy flower appreciated.

boy SBJ flower LOC good remain-PST.

A gloss often gives a reader more information about a given sentence, but is less comprehensible than a full translation. In presenting language data, using both a gloss and a translation is standard practice.

Again, as more terminology is needed, it will be introduced. Otherwise, as the Dothraki would say, *dothralates*: Let's ride!

Part I: Sounds

Introduction

When I was hired to create Dothraki for *Game of Thrones*, I didn't get many notes from the producers. Really there were only two things they wanted from the Dothraki language: (1) They wanted it to incorporate all the words George R. R. Martin had created in his books, and (2) they wanted it to sound harsh.

Harsh.

What does that mean when it comes to a language? Jason Momoa has described Dothraki as sounding like German. Many have described it as sounding like Arabic; a few like Russian. Most English speakers I've run into agree, though, that it sounds harsh. Why?

One tack might be to compare the languages that Dothraki is compared to. What sounds do they have in common? Quite a few, actually. Arabic, Russian and German have all of these sounds in common: *b, s, z, sh, k, l, m, n, a, i*… Quite a few more than that. Does having an *m* in a language make it sound harsh? Probably not. Almost every language on the planet has an *m* sound. In fact, a lot of languages feature those sounds listed above—including English—so perhaps we need to try something different. What sounds do Arabic, Russian and German have in common that English lacks? Turns out it's only one sound, and its phonetic transcription is [x]. You'll often see it spelled *kh* and it's often referred to as a "throaty" sound like the *ch* in German *Bach*. And, indeed, that does seem like a pretty "harsh" sound to an English speaker. It's so…clearing-your-throat-sound-y. That must be it! The presence of that sound makes a language harsh.

But hang on. You know what other languages have that sound (or if not that sound exactly, something very, very close to that sound)? Spanish and French. It's true. The *j* in the Spanish name *Javier* is pronounced pretty much exactly like the *ch* in German *Bach*. And the French *r* comes out as an even throatier version of this sound in words like *trente*, "thirty" and *produit*, "product". Do you know *any* English speaker that's ever described either French or Spanish as "harsh" or "throaty" or "guttural"?

Clearly there's more at work here than the presence or absence of a sound or two. In addition to the history of cultural stereotyping, which certainly plays a role here, it's the comparison of entire sound systems that produces a phonaesthetic character in the mind of the listener: the sounds present, the way they're combined, the intonational phrasing, and the rate of speech—plus a number of sociological factors. All of this is compared to the sound systems present in the mind of a speaker. So German may sound harsh to an American English speaker, but might not to a Dutch speaker from the Netherlands.

Even if the judgments are subjective, a conlanger can use the expectations of their users to achieve a particular phonaesthetic effect. In this chapter, you'll learn how to do just that.

For the most part, I will be focusing on how to construct naturalistic sound systems or **phonologies**. At the end, though, I will touch on sign language phonology and what you might do if you were constructing an alien sound system. As a general note, when you begin to create a language, you can start anywhere. I like to start with the sounds, though, so that's where we'll begin.

Phonetics

All spoken languages on Earth use a small subset of the possible sounds a human can produce. For example, an English word like *lava* is pretty understandable: you've got an *l* sound, a nice open *a* vowel, a *v* sound and then a kind of reduced *a* sound. Even though languages will use different sounds, it wouldn't be surprising to find that type of a word in any given language. It would be odd to find a word like *la§a* in a language. How do you pronounce *la§a*? Pretty much like *lava*, except instead of putting a *v* sound in between the *a* vowels, you clap your hands together once. So *l-a-CLAP!-a*. No natural language on Earth does this (except in songs or language games [remember that dog Bingo?]). There's no reason why a language couldn't do this (it'd be fairly simple to incorporate it into a language. Try replacing the sound *f* with a clap in English. Takes practice, but you can do it), it's simply the case that natural spoken languages don't.

Instead, oral languages utilize sounds made with the mouth, throat, tongue, nose and lungs. In this section I'll introduce you to those sounds, with examples from both natlangs and conlangs. First, though, I want to talk a bit about transcription.

In English, we use a number of sounds to convey meaning. Sometimes, though, the same sound will be pronounced differently, even if we're usually completely unaware of it. Try this test out yourself. Put your hand right in front of your mouth and pronounce the word *top*. Do it a few times. You should feel a nice puff of air on your palm. Now try pronouncing the word *stop*. Focus on the *t*. Notice anything different? Try saying the pair *top*/*stop* a few times. You should notice that that puff of air you feel when you pronounce *top* is totally absent when you pronounce *stop*. And, in fact, even though we

would call both of those *t* in English, the sounds are different. The *t* in *stop* is a regular *t* which we would transcribe this way: [t]. The *t* in *top*, though, is an aspirated *t*, and we would transcribe it like this: [tʰ]. That tiny little superscript ʰ tells you that the sound is pronounced with an accompany puff of air. When we write English, we don't bother noting the difference because English speakers don't distinguish the sounds. In phonetics, though, we do make a note of the difference, since even though it doesn't produce a meaningful distinction in English, it does in some languages.

Here, for example, are two different words of Hindi (ignore the funny tail on the *t*; we'll get to that):

टीक [ʈik] "teak"

ठीक [ʈʰik] "okay"

These two words differ only in one respect: the [ʈ] for the Hindi word for "okay" is aspirated (i.e. [ʈʰ]) and [ʈ] for the Hindi word for "teak" is not. So even though the distinction between [t] and [tʰ] isn't meaningful in English, such distinctions can be meaningful in other languages. This is probably why the two versions of *t* are spelled differently in Hindi (ट vs. ठ), but they aren't in English.

The spelling you see above in between brackets is what's known as **phonetic transcription**. Any time in this book that you see words in between brackets, it means that, more or less, this is *exactly* how the word is pronounced, and that the word will be written in the **International Phonetic Alphabet (IPA)**. This type of transcription contrasts with **phonemic transcription**, which gives you the most crucial phonetic

information of a word but leaves out some of the details. Below, for example, is the English word *top* in phonemic and phonetic transcription:

/tɑp/

[tʰɑp̚]

Notice that the phonemic transcription (which is always given between forward slashes) leaves off the aspiration marker ʰ. This is because it's considered to be a predictable pronunciation detail. The same isn't true of the aspiration and Hindi, which is not predictable. Here's the same transcription for the Hindi word for "okay":

/ʈʰik/

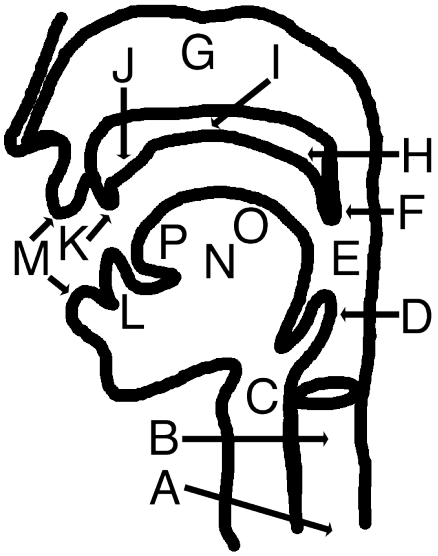
[ʈʰik]

They're the same, because the aspiration is a crucial factor in determining the meaning of the word in Hindi.

When you're creating your own language, you want to start with the phonetic level of detail. That is, you want to draw from *all* possible human sounds before narrowing it down and deciding which ones will be important for distinguishing meaning. I'll now take you on a tour of the plethora of sounds humans can make with their mouths.

Oral Physiology

Before we can talk about sounds, we have to talk a bit about what humans use to produce oral sounds. Below is what someone would look like if you sliced them in half from the top of their skull to their crotch (perhaps with a sharpened hat, à la Kung Lao from *Mortal Kombat*).



Each of those spots labeled above is referenced in the production of speech sounds. In order to produce a speech sound, in addition to the airstream, there is an **active articulator** and a **passive articulator**. The active articulator is the part of the mouth that moves to produce a constriction. The passive articulator is the part of the mouth that the active articulator either touches or gets near to in order to produce the consonant.

Below is a correspondence set of all the points in the mouth with the adjectival form of the word appearing in parentheses next to each noun (where necessary, further explication is given after the adjectival form):

A. Lungs (air coming from the lungs is referred to as *pulmonic*)

B. Larynx (laryngeal)

C. Glottis (glottal): At the bottom of the windpipe, the glottis is the space between the vocal folds.

D. Epiglottis (epiglottal): This is a little flap that closes when we swallow in order to ensure food goes into the stomach and not the lungs.

E. Pharynx (pharyngeal): This is essentially the back of the throat.

F. Uvula (uvular)

G. Nasal Cavity (sounds which feature air flowing through the nasal cavity are called *nasal*)

H. Velum (velar): The velum raises (i.e. closes) to prevent air from passing between the nose and lungs, and lowers (i.e. opens) to allow air to flow between the nose and lungs. When we breathe through our noses, the velum is lowered. (Note: Also referred to sometimes as the *soft palate*.)

I. Hard Palate (palatal)

J. Alveolar Ridge (sounds produced using the alveolar ridge as a passive articulator are referred to as *alveolar*)

K. Upper Teeth (sounds produced using the upper teeth as a passive articulator and the tongue as an active articulator are referred to as *dental*; sounds produced using the upper teeth as a passive articulator and the lower lip as an active articulator are referred to as *labiodental*)

L. Lower Teeth (dentolabial)

M. Lips (sounds involving the lips in some way are referred to as *labial*; sounds crucially involving both lips are referred to as *bilabial*)

N. Tongue (lingual)

O. Tongue blade (laminal): This is defined as the part of the tongue behind the tip.

P. Tongue tip (apical)

This may seem like a lot of terms to keep track of, but honestly, velar, alveolar, labial and palatal are the words you'll end up using the most. Feel free to bookmark this page, but I promise you by the time you get to page 90, you'll have this down (unless you decide to skip directly to page 90. That's cheating, and I guarantee nothing to a cheater).

Consonants

All sounds are produced either by having air pass out of the lungs and through the mouth, or by utilizing some sort of alternate air source that we can manipulate. Without force of some kind, no noise is produced (aside from the teeny tiny noises produced by having your organs move about). A **consonant** is a sound which puts some sort of obstruction in the way of the airflow, thereby changing the current and producing a different sound. Depending on how the current is affected, one can produce different types of sounds. The major divisions are listed below:

- Oral Stops: An oral stop is produced when airflow is interrupted completely (i.e. stopped), and the velum is raised, allowing no air to pass through the nose.
- Fricative: A fricative is produced when a tight constriction is formed somewhere in the mouth. Forcing the air through this tight constriction produces turbulent airflow which we interpret as different types of sounds.
- Affricate: A sound which begins as an oral stop but is released as a fricative.
- Nasal Stops: A nasal stop is produced when airflow is interrupted completely, and the velum is lowered, thus allowing air to pass through the nose.
- Approximant: An approximant occurs when the active articulator (the tongue or lips) approaches a position, but never forms a tight enough constriction to produce a fricative, resulting in a "liquid"-like sound.
- Flap/Tap: Referred to using either word, a flap or tap is when an active articulator is catapulted against a passive articulator.
- Trill: A trill is when the root of an active articular remains rigid allowing the tip to flap back and forth aggressively in the airstream.

- Lateral: A lateral is an L-like sound that allows air to pass around the sides of the tongue.

We'll examine each of these types of sounds in the sections below.

Oral Stops

Oral stops are amongst the simplest sounds humans use in language after language. Apart from other modification, a stop is actually characterized by the absence of noise. Here's what I mean by that. Trying holding an *s* sound (i.e. just say *s* and keep saying it). You probably sound a bit like a snake. Now try to do that same type of thing with *p*. If you actually tried this, what you probably did was sit there with your lips pursed and an awkward kind of pressure building up behind your lips. You'll notice that if you repeat this exercise with *k* and *t*, they all sound exactly the same—that is, like absolutely nothing.

That's what a stop is. In fact, what we hear when we hear a stop is the transition: moving from a vowel or some other consonant to a stop, and/or moving from a stop to a vowel or some other consonant. The distinction between the sounds produced when exiting or leaving a stop is what makes a *p* sound different from a *t*.

Here's a table which lists some of the more common stops found crosslinguistically:

Labial	**Dental**	**Alveolar**	**Retroflex**	**Palatal**	**Velar**	**Uvular**	**Glottal**
p, b	t̪, d̪	t, d	ʈ, ɖ	c, ɟ	k, g	q, ɢ	ʔ

Since this is the first time we've seen it, let me tell you why most of these cells have two members. Speech sounds feature what in linguistics we call **voicing**. Sounds may be either **voiceless** (sometimes called **unvoiced**) or **voiced**. In each of the pairs above, the voiced sound appears on the right. A quick lexical example to show us the difference between the two are the English words *staple* and *stable*. These two words

differ in only one respect, and that's the voicing feature of the second stop in that word. In *staple*, we have a voiceless [p], whereas in *stable* we have a voiced [b].

If you can't pinpoint the precise difference between these two, I'd like you to perform another throat experiment. I'd like you to place your hand against your throat right in the middle just above where the Adam's apple is (this protrusion, known as the laryngeal prominence, is present in all humans; it's just larger in males). Now what I want you to do is hiss like a snake again, except this time alternate between an *s* sound and a *z* sound. Something like:

sssssssszzzzzzzzzsssssssszzzzzzzzzzsssss

If you did the above, you should notice a considerable difference. For the *s* sound, your throat should have been relatively still. During *z*, though, you should have noticed a remarkable vibration. This is because *z* is a voiced sound. In order to produce a voiced sound, the vocal folds vibrate as air passes through the glottis. When the vocal folds vibrate in tandem with the production of a speech sound, that speech sound is referred to as voiced.

Now look again at the chart, and compare [p] and [b] again, as well as [t] and [d] and [k] and [g] (these sounds are spelled the same way in English, so they should be familiar). Compare the state of your throat in producing first the voiceless consonants, and then the voiced ones. You should be able to detect the same difference. Since you're pronouncing stops, though, the difference won't be as noticeable as it was with the *s*/*z* experiment. Nevertheless, the distinction is enough to produce a noticeable difference. Many languages make use of this distinguish produce different words.

With that in mind, here are some words in English where the six English stops mentioned above are the distinguishing factor:

lap [læp] ~ *lab* [læb]

boot [but] ~ *booed* [bud]

muck [mʌk] ~ *mug* [mʌg]

I want you to notice that no matter how the word is spelled, the phonetic transcription remains consistent. Even though there's a spurious *c* in *muck*, there's only a [k] in the transcription. IPA is used to transcribe the sound of a word, regardless of its spelling. This is important to keep in mind when examining English, whose orthography was devised by a team of misanthropic, megalomaniacal cryptographers who distrusted and despised each other, and so sought to hide the meanings they were tasked with encoding by employing crude, arcane spellings that no one can explain. (*"Ha, ha! I shall spell 'could' with an ell! They will be powerless to stop me!"*)

Before moving on to the sounds not found in English, take a look at the glottal column. Notice anything odd? [ʔ] looks pretty lonely up there, doesn't he? (And he also looks a bit like a question mark, making that last sentence look…clownish.) That sound is referred to as a glottal stop, and we've actually got it in English. When we say "uh-oh!" in that kind of exaggerated, mawkish way we use with children who appreciate it less and less with each passing year, we actually pronounce a glottal stop. A broad transcription of the phrase might be something like [ʔʌʔo]. That little catch in your throat in between the "uh" and the "oh" is a glottal stop. You'll hear it in Cockney pronunciations of words like "bottle", or in a native pronunciation of *Hawai'i*. It's a good sound used in a lot of languages, but the reason there's no voiced version is that it's

produced by smacking the vocal folds together. The vocal folds are what need to be vibrating in order to produce voicing. Consequently, a voiced glottal stop is impossible.

Now let's look at some of the other sounds. First let's look at the dental stops, for which the IPA has a terrible written convention (there really ought to be separate symbols for these). Dental stops are pronounced with the tip of the tongue against the back of the upper teeth. The difference is slight, but noticeable. Try checking out the native pronunciation of the Spanish word *desperado* compared to its pronunciation in English. Aside from a whole host of other slight differences, the initial *d* is natively a dental stop, whereas English's is an alveolar stop.

Hindi actually contrasts dental and retroflex stops, and lacks alveolar stops. A retroflex stop is pronounced with the tongue tip bent *backwards* to touch the roof of the mouth (the hard palate). To the ear of a native Hindi speaker, the English alveolar stops sound so different from their native dental stops that English borrowings routinely are spelled—and pronounced—with retroflex stops. Take a look at the following borrowings from English to Hindi:

डाक्टर [ɖɑktər] "doctor"

टिकट [ʈɪkət̪] "ticket"

टेलीफोन [ʈelifon] "telephone"

Since languages rarely contrast dental and alveolar stops, dental stops tend to be written without the little tooth underneath it (called a bridge), and a note is simply made at the beginning that the stops are pronounced with a dental place of articulation. This is the case with Dothraki, where the word *dil*, "shiny", is pronounced similar to the

English word *deal*, but with the tongue tip placed *right* behind the teeth when pronouncing the *d*.

The palatal consonants [c] and [ɟ] aren't rare, but they are often confused for (or pronounced as) others sounds. For example, the two sounds are quite similar to the initial affricates of English *chuck* and *jazz*, respectively. Whereas these English sounds are pronounced with the tongue tip initially contacting the alveolar ridge, true [c] and [ɟ] are pronounced with the tongue blade contacting the hard palate. To understand the difference in pronunciation, try pronouncing the name *Chuck* first as you would in English, and then second with your tongue tip touching your *bottom* teeth. If you can manage that, you'll be pronouncing [c].

Even though there is often some fluidity between palatal stops and post-alveolar affricates, there are some languages that contrast the two—one being Hungarian. Consider the pairs below:

csak [tʃak] "only" ~ *tyúk* [cuːk] "hen"

dzsem [dʒem] "jam" ~ *gyenge* [ɟeŋge] "weak"

Now these aren't minimal pairs (i.e. two words that mean different things but differ only in one sound, like English *bog* and *dog*), but it illustrates that the sounds are contrasted in the same environment (i.e. at the beginning of a word), and so a fluent Hungarian speaker will need to be able to produce and tell the difference between both sounds. When we get to the section where you're designing your own phonology, remember examples like this one, as it's always important to keep in mind what's *possible* and what's *likely*.

The last set of stops we'll discuss here are the uvular stops. Uvular stops are extremely alien to English speakers, but they're not that odd, crosslinguistically. You produce them by touching the back of your tongue to your uvula. The voiceless uvular stop is one many people will accidentally produce when yawning (and I mean a *big* yawn, where you could navigate a cruise liner into your mouth and still manage to fit a nice, fresh, crunchy carrot on top of that). To pronounce, open your mouth *as wide as you possibly can* (it should almost hurt). Now press your tongue down on the bottom of your mouth. Make it *as flat as you possibly can*. Now try to pronounce a [k] sound. If you're really keeping your tongue flat and you're managing to produce a [k]-like sound, you're successfully pronouncing the sound transcribed as [q]. It's the sound in Arabic loan words or names like Iraq and Quraan. It's also the sound spelled with a *q* in Dothraki, High Valyrian and Astapori Valyrian. Here's an example where the two contrast in Dothraki:

hake [hake] "famous"

haqe [haqe] "tired"

The sound is fairly common, crosslinguistically. It's found in most varieties of Arabic, but also in the various languages of the Inuit people, the Caucasian languages, and Mongolian. It's a good, strong, sturdy sound, and we shouldn't shy away from it.

The voiced counterpart to [q], however, is not nearly as common. We transcribe it [ɢ], and while it is found in plenty of natural languages, it often appears as a phonological or stylistic variant of some other sound (usually [q] or [g]). That said, in creating your own language, you can always make use of the contrast. Here, for example, is a minimal pair from Brent Scarcliff's Ancestral language:

[ɋa] "star fire"

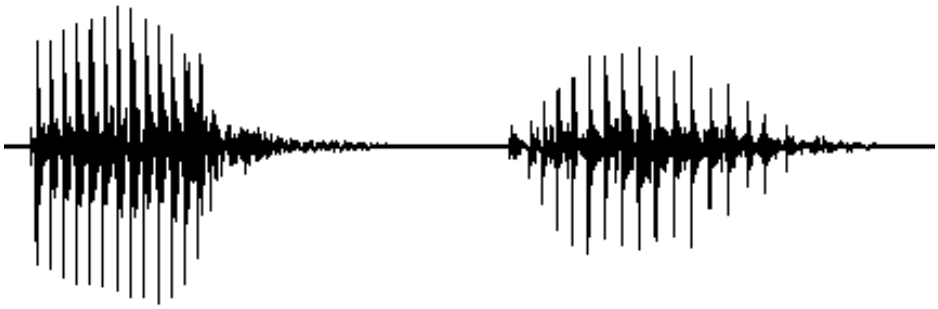
[ɕa] "blue-green algae"

There are other stops that exist (for example, the epiglottal stop [ʔ]), but they are vanishingly rare, crosslinguistically. Now we're going to move on to my favorites: fricatives!

Fricatives

Fricatives are the best thing since liberty. We produce fricatives by taking one of our active articulators (the tongue or lips) and putting them near one of our passive articulators and then just blowing out air. Unlike stops, we can hold fricatives as long as we've got air to blow out our lungs (cf. the *s*/*z* example from the Oral Stops section). There are more different types of fricatives than there are of any other type of consonant, and this is a good thing. Many languages don't take advantage of the plethora of fricative sounds a human being can produce, and this leaves me hollow inside. English is one of these languages. It is troubling.

To demonstrate the difference between stops and fricatives, though, take a look at these wave forms. First take a look at a wave form of me pronouncing the made up word [ata] (you can ignore the vowels; it's the [t] that's important):



All right, you're going to see three areas that are of interest. On the left and right you'll see two large areas that look a bit like dead trees reflected in a still pond on a windless day (or a still lake, actually. For scale, that'd be a pretty big body of water—at least to accommodate the size of tree I imagine). Those are the [a] vowels. Then there will be a region in the middle that looks like flat land. That middle region is the [t]. As you can see, the vowels are big and noisy, the first one tapers off (as you move through time

from left to right) and eventually culminates in silence. Then the new vowel pops up and tries to be as loud as the first one, but doesn't make it, since I put main stress on the first vowel, and partly how we realize that in English is to give extra volume to the stressed vowel (plus you lose air as you speak, so things will naturally quiet down unless you add greater force to your speech as you tail off). Thus, in order, you get [a], [t] and another [a].

Now take a look at [asa]:

Yowza! See how much more exciting that is? You still see the big foresty regions that are the vowels on either side, but now instead of tapering off into silence in the middle, there's a whole bunch of underbrush (imagine a swamp and those tall grasses and things that probably have real names but I'm not a botanist so leave me be). That rough region is the fricative [s]. Fricatives are much more rambunctious than any other sound (and, in fact, [s] is probably the most rambunctious of the bunch). They're steady, unlike trills, but the sound they give off is as unmannerly as a waterfall.

Here's a table of some of the most common fricatives, and how they're spelled in IPA:

Bilabial	**Labio-Dental**	**Dental**	**Alveolar**	**Post-Alveolar**	**Retroflex**
ɸ, β	f, v	θ, ð	s, z	ʃ, ʒ	ʂ, ʐ
Alveolo-Palatal	**Palatal**	**Velar**	**Uvular**	**Pharyngeal**	**Glottal**
ɕ, ʑ	ç, ʝ	x, ɣ	χ, ʁ	ħ, ʕ	h, ɦ

First, the following IPA symbols are identical to their most common realizations in English:

[s] *s* as in *sound* [sawnd]

[z] *z* as in *zoo* [zuː]

[f] *f* as in *fun* [fʌn]

[v] *v* as in *victory* [vɪktə˞ɹi]

The following sounds are also English sounds, but may be represented in a number of ways. As with any IPA character, the goal is to focus on the *sound*, not the *spelling*:

[θ] *th* as in *thigh* [θaj]

[ð] *th* as in *thy* [ðaj]

[ʃ] *sh* as in *sheep* [ʃip]

[ʒ] *g* as in *genre* [ʒɑnɹə]

Couple of notes on these. Notice that English actually does distinguish the sounds [θ] and [ð]; it just doesn't do a very good job of it. We spell both *th*, and it's up to

the speaker to realized that the pronunciations of *th* in *this* and *thin* are totally different ([ð] and [θ], respectively). I consider it to be most unmannerly (which, I guess, is what you'd expect from a wily, unmanageable pair of fricatives). Consider Arabic, which also has both of these sounds but which spells them differently:

ثاب [θaːba] "to come (e.g. to one's senses)"

ذاب [ðaːba] "to melt, to dissolve"

See? Now *that* makes sense. And don't get me started on the world's most perfect sound [ʒ], from English *genre* and *measure* and *mirage*. What an awesome sound! And I think it's a sound English speakers have a fairly good grasp of (try pronouncing *genre* with the *j* sound, [dʒ]. If that won't convince you that [ʒ] occupies its own unique space in English, nothing will). Yet ask an English speaker to spell it, and we throw up our hands. *G* sometimes? *J* like never, but we think sometimes it might be anyway? *S* sometimes, like in *vision*? We've got nothing. Hungarian spells it *zs* (hence Zsa Zsa Gabor), and Arabic spells it with a ج (at least for dialects that have the [ʒ] sound, as opposed to something else, like [g] in Egyptian Arabic), but English is stuck with an ensemble cast of C-list spelling conventions for an Oscar-winning sound. Pitiful.

There are two remaining sounds that occur in English, though one in only a marginal way. The [h] you see up there is a regular *h* as you would expect in a word like *help*. The one with the combover sitting next to him, though, has the same relationship to [h] as [z] does to [s]. We sometimes use this [ɦ] sound in pronouncing words like *ahead*. The vowels on either side of the *h* are voiced, so sometimes English speakers will voice the *h* and pronounce it as [ɦ]. That's the best I can do for you there. I don't think there's a language that exists that contrasts [h] and [ɦ] regularly, and the status of

[ɦ] itself is somewhat nebulous (is it a fricative? Is it even a consonant? Is it just something that happens to certain vowels?), but plenty of languages make use of it as either a stylistic or phonemic variant, so it's a good one to stuff in your conlanging duffle.

The rest of the sounds on the chart can be explained in terms of other sounds. For example, [ɸ] and [β] are just like [f] and [v], except that you pronounce them by pursing *both* of your lips and blowing. The sound [ɸ] is literally identical to blowing on hot soup or coffee. [β] is just the voiced version of that sound. The phonetic result is that both [ɸ] and [β] sound a bit less rambunctious than [f] and [v]. This is partly why the bilabial fricatives are rarer than the labio-dental fricatives. Nevertheless, they do occur in several languages, and some even make a contrast between them—the most famous example being the West African language Ewe. There's a minimal quadruplet where each sound is contrasted:

efe [efe] "nail" or "debt"

eɸe [eɸe] "year"

eve [eve] "two"

eʋe [eβe] "the Ewe people"

According to Ian Maddieson, a linguist from UC Berkeley, a lower F1 (we'll talk about formants when we discuss vowels) for the bilabial fricatives and a striking visual contrast between the labio-dental and bilabial fricatives may aid speakers in distinguishing the sounds which, otherwise, are quite similar acoustically. In other words, face-to-face communication is key in keeping this distinction alive, so if you're creating a language for people that only communicate via cell phone, this may not be a good distinction to build in.

The retroflex and alveolo-palatal fricatives can be thought of as different sounding versions of their post-alveolar fricative counterparts. In other words, they're different *sh* sounds. For example, whenever anything is spelled *sh* in Japanese rōmaji, the sound is not [ʃ], as it is in English, but [ɕ]. The same distinction holds for the sound sometimes romanized as *j* in certain rōmaji systems, where English has [ʒ] but Japanese has [ʑ]. The difference is that while with English [ʃ] you position the tongue tip behind the alveolar ridge, with Japanese [ɕ] the blade of the tongue is moved up towards the hard palate, with the tip of the tongue pointing towards the lower teeth. The best way to perfect the sound is to watch anime. *Lots* of it. Aside from anything coming out of Studio Ghibli, I recommend *One Piece*, or *Trigun*, or *Paranoia Agent*, or *Samurai Champloo*, or *His and Her Circumstances*… Lot of good stuff, actually. Dang, can I switch topics? I want to talk about anime now…

The retroflex sounds are produced by bending the tongue backwards and making a *sh* sound (or a sound like [ʒ] for the voiced variant). If you round your lips, you'll be producing a sound that I commonly associated with older American English speakers from the Mid-Atlantic region. It's also a sound found in Hindi, among many others. Both the voiceless [ʂ] and the voiced [ʐ] are quite common, crosslinguistically, either as the solitary *sh*-like fricatives, or contrasting with other *sh*-like sounds.

The remaining fricatives can all be described as *h*-like. This makes the voiceless variants easy to describe; the voiced variants will require a bit more work. For that reason, we'll start with the voiceless versions. First, I want you to compare your pronunciation of the following two English words: *he* and *ha*. Focus on the *h*. Notice a difference between those two? For *he*, the tongue body is raised quite a bit, whereas for

ha it's probably at the bottom of your mouth. In fact, in English the *h* in *he* is [ç], while it remains [h] in *ha*. An English speaker will think of those sounds as being the same, but they're really quite different.

Keep a pin in that one, because now we're going to talk about the "harsh" sounds. If you're an American, think about that sound you make when someone pulls out a glorious pun like, "'I see', said the blind man as he picked up his hammer and saw." It's usually accompanied by a roll of the eyes. We tend to write it something like *ach*. That sound is [x]. That's a voiceless velar fricative, and, basically, *t* is to *s* as *k* is to this sound. If you go to pronounce a [k], but you release it *slowly*, you'll be releasing into the sound [x].

Okay, got that? Now do it farther back. It should still sound throaty and white-noise-y (if you're young enough to be unfamiliar with white noise on a television, it's the sound they play between sketches on *Robot Chicken*), but it should be moving closer to triggering your gag reflex. That is the sound [χ]. If you're pretty good at speaking French like a native, this is the sound its *r* makes in words like *trop*. Most languages don't draw a regular distinction between [x] and [χ], but those that do, look out! Seriously, go check out the phonemic inventories of Ubykh, which has seven [x]/[χ] sounds, and Tlingit, which has eight.

The last voiceless phoneme may seem like one of the most alien, but it's actually the easiest to explain. Pretend you're fogging up a mirror. If you try to do that by pronouncing an [h], you'll get nothing. If you *really* want to fog up that mirror, you will naturally produce [ħ]. It's a wonderful sound that's present in Arabic and a number of other languages. And unlike word-final [h], which has a tendency to disappear in

language after language, [ħ] is nice and sturdy. To make a tortured weather-based comparison, if [h] is a light breeze, [ħ] is an F5 tornado of hot, breathy *h*-ness, and I love it.

Now for the other halves. Yes, technically if you just vibrate your vocal folds, you'll be able to produce the voiced versions of [ç], [x], [χ] and [ħ], but the truth is a lot messier. If you look at a full IPA chart, the voiced versions of each of these sounds either will be described as straddling the line between fricative and approximant, or will have an approximant sound with a separate character right below the voiced version. This isn't an accident. Many natural languages that have the voiced versions of these non-strident fricatives (i.e. [ʝ], [ɣ], [ʁ] and [ʕ]) will use them as line-straddling fricative/approximants. This is the case with [ɣ] in Spanish ([β] too), [ʝ] in Greek and Hungarian, [ʁ] in French and German, and [ʕ] in Arabic. It is perfectly possible to pronounce each one of these as an honest-to-goodness voiced fricative, but a combination of factors (maintenance of voicing plus the necessity of maintaining a tight constriction in places where it's not as easy to do so) contributes to these voiced fricatives becoming approximant-like. This is something to file away in your conlanging duffle as you think about using these odd (but very cool) sounds in a conlang. I'll talk a bit more about these in the section on approximants.

Now that we've discussed fricatives, the remaining sections will be remarkably shorter. Oral stops and fricatives are the most diverse sounds humans employ in spoken language. There's a lot of variety in what's left taken as a whole, but within individual manners, there's a lot less diversity than there is amongst fricatives and oral stops.

Affricates

An affricate is a stop followed by a fricative in the same place of articulation. It might seem odd to give a term to two consecutive sounds, but in language after language affricates tend to be thought of as single sounds. You'll see why when you seem some examples. Here are some of the more common affricates found in the world's languages:

Labial	**Alveolar**	**Post-Alveolar**	**Alveolo-Palatal**	**Retroflex**	**Palatal**	**Velar**	**Uvular**
pf, bv	ts, dz	tʃ, dʒ	tɕ, dʑ	ʈʂ, ɖʐ	cç, ɟʝ	kx, gɣ	qχ, ɢʁ

Before discussing the rest, I want to draw your attention to [tʃ] and [dʒ]. These sounds are spelled *ch* and *j* in English, respectively. Hopefully this will help to demonstrate how we can think of consecutive sounds as single sounds. Here are just two words that illustrate the distinction:

chump [tʃʌmp]

jump [dʒʌmp]

Technically the first sound should be aspirated, but we'll leave that side for the moment. The point is that English speakers think of these as single sounds. The same tends to be true of the rest of these, for languages in which they occur as single sounds. For example, we have [dz] in words like *kids* [kɪdz], but we don't think of the [dz] as a single sound. In Albanian, though, it's quite comfortably its own sound. The word *xanxë* [dzandzə], for example, has two instances of [dz] (spelled *x*).

With that in mind, the pronunciation of each of the affricates listed in the table above can be derived from combining the pronunciation of the stop with the

pronunciation of the fricative. Just two notes on some aberrant conventions, though. Technically for [tʃ] and [dʒ] a homorganic stop would be a post-alveolar stop, meaning that the [t] and [d] should be transcribed [t̠] and [d̠], respectively. In practice, this is much more trouble than its worth, as post-alveolar stops never contrast with alveolar stops in a meaningful way in any natural language. The same can be said for the affricates [pf] and [bv], which should properly be [p̪f] and [b̪v], respectively. And while we're on the labial section, both [pɸ] and [bβ] are perfectly possible; they're just not as common as [pf] and [bv]. The same goes for other less common affricates, such as [t̪θ] and [d̪ð].

Finally, a bit of a typographical note. Affricates ought properly to be written with a tie bar—i.e. [t͡ʃ] instead of [tʃ]. I've not done so here (and will remain consistent with this practice) as the symbols come out clearer, and it simply needs to be noted whether or not a stop followed by a fricative is to be considered an affricate in the language in question.

Nasal Stops

Nasal stops (also called nasals) are basically oral stops where air passes through the nasal cavity and out the nostrils by means of a lowered velum. You can these of these as *n* and *m* sounds. To experience the difference oral and nasal stops, you can do a number of things. But your finger directly under both nostrils, as if you're pretending to give yourself a fake mustache. Now pronounce, in succession, the English words *dud* and *nun*. You should notice that there's some warm air being blown on your fingerstache for *nun*, but not for *dud*. The effect is even more pronounced if instead of a finger you hold a mirror under your nostrils. It'll fog up nice for *nun*. You should also notice a marked difference if you pinch your nostrils and try to pronounce *nun*. *Dud*, on the other hand, should work just fine.

The fact that the velum is lowered has consequences on the types of nasal sounds produced. Let's start with *m*, for example. What makes an *m* work is the fact that your lips are closed, but air continues to pass through your nostrils via your nasal cavity. Airflow is completely stopped going out of the mouth (this is why they're called nasal *stops*), but air continues to flow through the nose. Keep this in mind as I show you a table of all possible nasal stops:

Bilabial	**Labio-Dental**	**Dental**	**Alveolar**	**Retroflex**	**Palatal**	**Velar**	**Uvular**
m	ɱ	n̪	n	ɳ	ɲ	ŋ	ɴ

Notice that the uvular nasal stop is the last stop moving from the lips backwards in the mouth. This is why. Remember that a nasal stop *simultaneously* (not consecutively) requires: (1) oral airflow to be stopped; (2) nasal airflow to be continuous

(i.e. a lowered velum). Now let's imagine a glottal nasal. How would that need to work? The velum would have to be lowered, of course. That's not a problem. Airflow in the mouth would need to be stopped. That's also not a problem: airflow is stopped at the glottis. Finally, air would need to pass continuously through the nose while airflow is stopped through the mouth. That's a problem. Since the air flows out of the lungs, it's stopped where the stop closure is. If that stop closure is the glottis, air can't pass through to exit the nasal during the stop closure. Consequently, a glottal nasal stop is literally impossible.

Now let's take a look at the uvular nasal [ɴ], though. The back of the tongue will be touching the uvula, which is just the punching back hanging off the end of the velum. This will stop air from flowing through the lungs and out the mouth. Since that's the very tip of the velum, though, the velum can still be lowered, and air can still pass through the nose while air is prevented from exiting the mouth. This is why the uvular nasal [ɴ] is the last possible nasal stop.

Most nasal consonants are fairly easy to understand and pronounce. For example, depending on your dialect, all but the last one are found in English. Here's where you'd see them:

mom [mɑm]

invite [ɪm̩vajt]

month [mʌn̪θ]

next [nɛkst]

burner [bə˞ɳə˞]

minion [mɪɲjʌn]

tongue [tʌŋ]

I will admit, a couple of those are dicey, but if you have an *r*-ful dialect of English (so not New York, not Britain), try pronouncing *burner* and *winner* and see if the *n*'s feel different. They do for me, and I'm all the way in kooky California, where everything's made up and the points don't matter. I bet that nasal's *crazy* retroflex in West Virginia.

One thing you should notice if you look at *invite*, *month* and *minion* is that what we write as *n* tends to take on the characteristics of what follows it. This is something that's extremely common crosslinguistically, but not universally so—and not even in the same language. Notice that we have the same *in-* in *indecent* and *improper*, but the *n* doesn't become an *m* in *unpopular* (either in spelling or pronunciation). Most of the time, though, many of these nasals will show up when they're next to other sounds. In English, the only ones that are truly free to appear anywhere are [m] and [n]. The next most free is [ŋ], but notice that we can't do it at the beginning of a word. This is why last name's like Nguyen give English speakers fits. Some English speakers will allow [ɲ] in initial position (e.g. certain speakers' pronunciation of *news* [ɲuz]), but none can deal with it in final position. This is why we pronounce *champagne* [ʃæmpen], with final [n], rather than something closer to the French [ʃɑ̃paɲ], which has a true palatal [ɲ] in final position.

Something else to notice in this table is that, unlike in the previous three, there's only a single entry for each place of articulation. This isn't an accident. Crosslinguistically, nasals are voiced by default, and we consider the voiced versions to be the basic form of the nasal. This doesn't mean that voiceless nasals are impossible, though—far from it! The symbol for any voiceless nasal is the same as its voiced

counterpart with a little circle underneath it (or above it, if there's no room). So a voiceless alveolar nasal looks like this: [n̥]. To pronounce them, just breathe through your nose. If you breathe out of your nose with your lips closed, that's a voiceless bilabial nasal. Putting it in between vowels is what makes it a speech sound. And while they're crosslinguistically rare, voiceless nasals are found in a variety of languages—most of the time simply as contextual variants of their voiced counterparts, but not always. Sometimes they distinguish meaning (in Burmese, for example). It's simply uncommon.

Approximants

Approximant is a catch-all term for sounds which aren't really fricatives, but also aren't trills or laterals. Most approximants are very closely tied to a vowel sound, to the point that approximants are sometimes written with vowels characters (indeed, one alternate name for an approximant is a semi-vowel). To understand the nature of approximants, I want you to try slowing down your pronunciation of the *y* consonant in *yawn*. Slow it down as much as you possibly can and see what happens. If you slow it down enough and hold it, you'll eventually be pronouncing the *ee* vowel from *need*, which we transcribe [i]. In effect, when [i] is turned into a consonant, it becomes the *y* in *yawn.* The same holds for the other approximants you'll see below, and the same strategy is used to produce them. Essentially, you take the vowel and push it a little bit further and then release it, giving you the liquid sound that is an approximant.

It may be a good idea to revisit this section after we've discussed vowels. Bearing that in mind, below is a list of the most common approximants found crosslinguistically, and they're associated vowel:

Labio-Palatal	**Labio-Velar**	**Alveolar**	**Retroflex**	**Palatal**	**Velar**	**Pharyngeal**
ɥ	w	ɹ	ɻ	j	ɰ	ʕ
y	u	ɜ˞	ɚ˞	i	ɯ	ɑ

The idea with the associated vowels is that if you pronounce those vowels and push them further, as with the *y*/*ee* experiment, you'll get the approximant version. Even so, those associated vowels are just approximations (note: puns are the lifeblood of linguistics). The two that are easiest to understand are [w] and [j], which are both found

in English in *watt* [wɑt] and *yippy* [jɪpi], respectively. The rest will need a little explanation—first and foremost the upside-down *r*-looking thingies.

We treat the language we grow up with as simple, logical, basic—the way language *should* be. Other languages are always difficult and bizarre. The judgments are, of course, based on what we understand language to be based on our own patterns, but some things can be measured at least somewhat objectively. One of them is how difficult it is to pronounce a sound. Turns out one of the most difficult sounds found in the languages of the world is the English *r*. Yes, we have bizarre consonant clusters (*crafts*? Really?), and make minute vowel distinctions speakers of other languages find difficult to hear and reproduce (Spanish speakers always have to be careful when they pronounce *beach*!), but our *r* is, hands down, the most awful thing in the entire language. Most of the time (and I will do so here, as well), you'll see it transcribed as [ɹ]. Truth is it's probably closer to [ɻ]—the retroflex approximant—but it's also accompanied by rounded lips. And to make matters worse, it's not even fully retroflex; the tongue tip kind of points forward a bit. *For some speakers*. (Because why would all English speakers pronounce *r* the same when literally every speaker can pronounce it slightly differently? [Hear that? That's the sound of us trolling anyone who has to learn to speak English.])

If you put a spork to my head and forced me to transcribe my own pronunciation of *r* as narrowly as possible, the monstrosity I'd probably feel comfortable producing is this: [ɻ̟ʷ]. It doesn't even render properly. It's hideous.

So when you're see something like [ɹ] or [ɻ] in the description of a language, it's best to hear an example first—and the same goes for using it in your language.

By comparison, the rest of the approximants are simple. For example, the crazy looking [ɯ̟] is pronounced just like [w] but with the lips completely unrounded. You may find this pronunciation in certain varieties of Spanish for *g* in between vowels, as with *lago* [laɯ̟o], "lake". The one that looks like this with one fewer bump—[ɥ]—is pronounced just like [j], but with completely rounded lips. For those who've studied French, this is the sound that begins the word for eight, *huit* [ɥit].

Our last approximant is [ʕ], which we saw before in the section on fricatives. Most often in the languages that it's found in it is an approximant rather than a fricative, and that's the case with Arabic. To pronounce, just pronounce a nice long *aaaaaahhhhh*, like you've got a tongue depressor in your mouth, and then push your tongue further back, till it feels like you're going to choke on your own tongue. *That's* the sound. Homer Simpson approximates it fairly well when he sees food he's really excited about eating. It takes a bit to train your brain to think of this sound as a speech sound, but once you get it, it's actually pretty manageable—and a fun twist to throw into a language that needs a little something.

There are other approximants (potentially an infinite amount, if you consider the definition I gave in the first paragraph), but these are the major ones. Others will be represented with fricatives, as I discussed in the fricatives section. Some straddle the boundary between fricative and glide, such as [ʍ], which is the old fashioned pronunciation of English *wh* (think of Stewie and Brian in that Cool Whip commercial). Another that managed to get its own symbol was [ʋ], which is like a [w] but instead of being pronounced with lips, it's pronounced with the lower lip and upper teeth, like a [v].

Flaps, Trills and Laterals

The last set of consonants will be discussed together. All of these are basically different types of *r*'s and *l*'s. First we'll start with the *r*'s: the flaps and trills. Here's a list of the most common flaps and all possible trills:

Manner	Bilabial	Alveolar	Retroflex	Uvular	Epiglottal
Flap		ɾ	ɽ		
Trill	ʙ	r	ɽr	ʀ	ʜ

Let's start with the simple one. The symbol [ɾ]—an orthographic *r* without the hook on the upper left-hand side—stands for a sound that is ubiquitous in American English. If you're an American English speaker, compare the difference between the *tt* and the *t* in *attitude*. In addition to the puff of air accompanying the *t*, the *tt* should sound quite a bit different. In fact, most people will pronounce *matter* and *madder* the same (if you're from Britain or certain parts of Canada, this may not hold true for you, but it does for the majority of American English speakers). The same goes for similar pairs (*utter*/*udder*, *mutter*/*mudder* [type of horse], *metal*/*medal*/*meddle*, *bitter*/*bidder*, etc.). If asked to describe the sound, most English speakers would probably say it's a *d* sound, but really, it's quite different from *d*. Compare the *dd* in *addition* (a true [d]) to the *dd* in *additive*. The latter is much, much quicker. In fact, the sound is pronounced identically to a single *r* in the same context in Spanish. Despite its historical origin as either *t* or *d*, this sound is [ɾ] in American English.

I want you to keep this sound in mind as we move on to the trilled [r]. This is the double *rr* of Spanish that gives some English speakers trouble. Some will say they can't

pronounce it. This is not so. If there's nothing wrong with your speech organs (your tongue is fully formed and flexible and you have an intact alveolar ridge), you can pronounce the trilled [r]. You're basically doing a solitary version of it everyday when you pronounce [ɾ] in American English (this isn't exactly true, but it's close enough). In fact, one of the places where I think English speakers come closest to producing the trill naturally is in karaoke with the Red Hot Chili Peppers song "Give It Away"—specifically in the chorus. The chorus—"Give it away / Give it away / Give it away now"—is really quick, so if you want to keep up, the flap in *it* can—and often does—become a short trill. To illustrate, here's how the first line of the chorus would be transcribed in normal speech:

[gɪv ɪt əwej]

Now here it is when sung quickly (notice it basically becomes a single word):

[gɪvɪɾəwej]

And when sung very quickly and not carefully, it can become this:

[gɪvɪɾəwe]

Granted, the singer's got to really get into the song before they'll naturally produce the last example, but it's there. It's a possibility for English speakers, and I think that a true trill is available to all, whether you can hold it for a fraction of a second or ten seconds.

The thing that makes a trill work is the root of the articulator (the tongue, in the case of [r]) remains strong, and the tip goes limp. This allows the air to flutter through the tip, forcing it to oscillate back and forth between the point of articulation and the open air. Each flap is a period, allowing one to produce longer or shorter trills. If you go up to try to pronounce a trill and you just get air (a truly deflating feeling, I know; we've

all done it), it probably means something went wrong in the description above. Either you didn't hit the alveolar ridge by the time the air got there, the tip of the tongue was too stiff, or the root of the tongue was not strong enough. More often than not it's the latter. That tongue root has to be ready to go into *battle* when it's going to pronounce an alveolar trill. It's a soldier, and the tongue tip is its flag. Let it fly!

The same story applies to the retroflex sounds. In fact we do find the retroflex flap in English. If an [ɻ] precedes or follows a *t* or *d* that would ordinarily be flapped, the result is [ɽ] instead of [ɾ]. So, for example, in words like *water*, the *t* is [ɽ] rather than [ɾ]. The retroflex trill is a bit rarer, but the same rules apply for pronouncing it—just with the tongue bent back a little bit at the start. This is a rare sound, so I wouldn't worry about it too much, or use it in a conlang without clear motivation.

Another rare sound that is nevertheless quite well known to most speakers is the bilabial trill [ʙ]. Some linguists have referred to this as the "raspberry" sound that we'll make with our lips, but I don't know what kind of raspberries they're doing. To me, the raspberry is put your lips against someone's stomach and blow out air, making a kind of trumpet noise, and delighting certain types of children. This isn't a spitting sound: it's a rolling, rumbling sound. It's the sound horses make with their lips and that we frequently imitate for no sane reason. Singers will be familiar with the sound as it's often employed in vocal warmups. It's actually used in natural languages, but it's vanishingly rare. Larry Hyman, a famous Berkeley linguist, once demonstrated two words from a language I can't remember, but I'm certain that the word for *goat* in this language was [ʙɨ]. This delighted me.

The final two trills are common, but difficult to pronounce. [ʀ] is one of the ways to pronounce the German and French *r*. It was once the most common way to pronounce *r* in French, but it's become old fashioned. Several dialects of German, though, do still use a trilled [ʀ] for their realization of *r*. It's very, very similar to what one does with one's throat when gargling. It's a deep, resonant, throaty sound, similar to a cat purr. To pronounce it correctly, though, it should roll. Failed attempts will produce a uvular fricative, and the result will sound harsher than the true trill.

The last symbol, [ʜ], is used for the epiglottal trill, which is pronounced even further back in the throat. If you were trying to do the uvular trill and you felt like you were almost going to choke (or it sounded to others like you were), you were probably pronouncing the epiglottal trill. It's not as common as [ʀ], but it's still relatively common, crosslinguistically.

Incidentally, voicing itself is a trill. To produce voicing, you keep the glottis still and force air through your larynx, causing your vocal folds to flap—and that is what produces voicing. Aside from that, there is one more trill the body can produce, but it's impolite to discuss it (and don't ask a phonetician to demonstrate it).

Flaps and trills can be voiced or voiceless, but they are most commonly voiced. If they are devoiced, it's usually when they occur next to voiceless sounds, but certain languages have used a full-fledged voiceless alveolar trill to distinguish meaning. It's not common, though.

Now onto the *l*'s. Laterals are sounds where air passes around the tongue, giving the consonant its characteristic *l* sound. Here are some common lateral sounds:

Manner	Alveolar	Retroflex	Palatal	Velar
Approximant/ Flap	l ɾ	ɭ	ʎ	ʟ
Fricatives/ Affricates	ɬ, ɮ tɬ, dɮ			

You can consider the alveolar [l] to be the "normal" *l*. Turns out, though, the English *l* is a bit more special—so special that one of its more common variants is called the Dark L by linguists. A true [l] is the type you'll have in words like *love* [lʌv]. Many languages will have this sound for its *l* sound, but on occasion you'll find the retroflex [ɭ], which is pronounced similarly but with the tongue tip bent backwards. The palatal [ʎ] is one that occurs in English before [j], e.g. in words like *billion* [bɪʎjʌn]. It occurs as a sound in its own right in many, many languages, though. In Italian, for example, it's the sound that is typically spelled *gl*, and it does serve to contrast meaning. Here are a couple examples:

li [li] "them (accusative, third person masculine plural clitic pronoun)"

gli [ʎi] "the (masculine plural definite article)"

The velar [ʟ] is another sound that makes you feel or sound like you're choking. To make it, try planting the tip of your tongue firmly beneath your lower teeth and pronouncing any kind of *l* you can. Most likely the result will be [ʟ]. It's a fairly rare sound, and probably one to be avoided. A much more common sound is the *velarized* alveolar lateral, which looks like this: [ɬ]. This is the English Dark L sound, and it's quite common. Compare the *l* sound in *love* to the *l* sound in *filth*. Notice how the second one sounds swallowed? That's the [ɬ] sound. You'll find it pretty much any time the English *l* doesn't occur before a vowel (e.g. *thrill* [θɹɪɬ], *girl* [gə˞ɬ], *felt* [fɛɬt], etc.), and when it

occurs before [ɚ] (e.g. in *sailor* [seɬɚ]. It sounds very similar to the velar [ʟ], but the difference is that the tip of the tongue does (or can) touch the alveolar ridge in [ɬ]. Thus, the tongue has the position of [l], but the action of the sound is velar—a kind of combination of [l] and [ʟ].

The sound [ɻ] is how *r* is pronounced in Japanese. It's like [l], but pronounced much quicker. If you try to do an [l] as fast as you possibly can, you should be able to approximate this sound, though it may actually sound a bit like an odd kind of *r* to you. This is one of the things that makes Japanese so difficult for an English speaker to process and pronounce, but the sound can be achieved with practice.

The last set of sounds are the infamous lateral fricatives and affricates. The voiceless and voiced lateral fricatives are found in Welsh and Mongolian, respectively. Here are two examples:

lladd [ɬɑːð] "kill"

лам [ɮam] "monk"

The voiceless lateral fricative is fairly easy to articulate. Try pronouncing [h] followed by [l] (kind of like saying *hello* without an *e* vowel). If you can pronounce the sounds consecutively, try your best to merge them. That is the voiceless lateral fricative [ɬ]. It's a sounds that sounds like it *must* be a combination of two other sounds. For example, know the name Floyd? That actually came from English speakers mishearing the traditional Welsh name *Llwyd*—anglicized to Lloyd—which is not pronounced with an [l], as it is in English, but with [ɬ]. The sound was misheard by English speakers as [fl], and a new name was born.

The voiced version of this sound is, in my opinion, one of the most difficult sounds to produce. Using the same analogy as before, [s] is to [z] as [ɬ] is to [ɮ]. It sounds like a weird, gooey combination of [ɣ], [ʒ], [z] and [ɬ]. It's terrifying.

As for the fricatives, slap a [t] in front of [ɬ] and a [d] in front of [ɮ], and you've got it. The sound [tɬ] is now world famous thanks to its place in the Klingon language. If you haven't heard it before, it might surprise you to learn that there is no [k] in Klingon, because, well, there's a *k* right there in the name: *K*lingon. In fact Marc Okrand decided to take a cue from the fact that the word begins with a *kl* to totally reinterpret the sound. Instead of [kl], the word Klingon properly begins with [tɬ]. If you pronounce it right, the two sounds are surprisingly close. Try it out: [klɪŋɑn] vs. [tɬɪŋɑn]. It was a clever choice, which future *Star Trek* writers completely ignored, coining supposedly native Klingon words and names like *Kronos, Kruge, Kurn*, and *Kahless* (the pronunciation of which always amused me, since it's homophonous with *k-less*—which is what Klingon is). Okrand has been pretty good natured about such gaffes, which is probably for the best, since this treatment is par for the course if you want to work in Hollywood.

Non-Pulmonic Consonants

Everything so far has given you a fair representation of the consonant sounds of the world—let's say about 95%. What of that other 5%? Come behind the curtain and take a look.

A **non-pulmonic** consonant is one that doesn't use the pulmonic airstream mechanism. In other words, these sounds are *not* produced via air that comes out of your lungs. There are three main types of non-pulmonic consonants: **ejectives**, **implosives** and **clicks**. As a group, they're not very rare, but it is out of the ordinary to find one of them in a natural language.

Ejectives are consonants that are pronounced at a normal place of articulation while the glottis is closed. This may sound difficult, but I've got a really simple way to demonstrate it. To produce an ejective, hold your breath (and I mean *really* hold; don't allow any air to creep in or out). Got it? Now do your best to pronounce a [t] or a [p] or a [k]. If you managed it, you produced an ejective version of that consonant. It should sound quite percussive—as if you were going to try to kill someone across the room by pronouncing a [t] at them. For fun, find a male speaker with a particularly large Adam's Apple (and if you are such a male speaker, find a mirror). Once you've got one, have him not pronounce an ejective, but get ready to. In fact, have you get ready to pronounce an ejective, then leave off, then get ready to again, etc. His Adam's Apple will dance for you. Anyway, after that when he pronounces the ejective, the Adam's Apple will give a little upward jump, like a frightened kitten. It's great fun to watch.

The reason for this is that with your glottis closed and a stop formed at some place of articulation (let's say the alveolar ridge), the air is trapped. The ejective is

produced by causing the creepily unattached hyoid bone and glottis to raise, which increases the pressure inside that closed off region of the larynx. When the pressure is released from the front (i.e. the tongue leaves the alveolar ridge), that pressure explodes out the mouth, producing the percussive ejective.

There are as many ejectives as there are voiceless consonants above the glottis. To write an ejective, you place a little apostrophe after the symbol, like so: [pʼ], [tʼ], [tʃʼ], [sʼ], etc. An ejective is necessarily voiceless, when you consider how it's produced. In order to produce voicing, air must flow out of the lungs to vibrate the vocal folds. If air is closed off from the lungs via a closed glottis, the vocal folds will of necessity remain still, meaning that all ejectives will be voiceless. Crosslinguistically, [kʼ] is the most common ejective, with those pronounced further away from [kʼ] being rarer ([pʼ] is quite rare). And, in generall, stops and affricates are more common than fricatives, though fricatives are possible. If you'd like to try pronouncing a couple of them in a word, here are some contrasting stop options from an early conlang of mine called Njaama:

[topa] "ring"

[tʼopa] "name"

Next we'll discuss implosives: the anti-ejectives. They have their own symbols, which you can see here:

Bilabial	**Alveolar**	**Palatal**	**Velar**	**Uvular**
ɓ	ɗ	ʄ	ɠ	ʛ

Implosives are traditionally described as obligatorily voiced and obligatorily non-pulmonic. Turns out most implosives actually do involve a little pulmonic voicing—and that voiceless implosives are possible, though slightly different in manner—but it's best

to steer clear of that whole hornets' nest unless you plan to pursue a Ph.D. in linguistics with an emphasis in phonetics. What makes an implosive implosive is a rapid rush of air *into* the mouth in producing the sound. This is done by forming a seal for a stop and *lowering* the glottis, which results in negative pressure inside the mouth. When the stop is released, air rushes in, and it gives the stop consonant a very bizarre sound.

Ordinarily I'm pretty good at coming with examples of sounds that occur in one way or another in English, but for this, the best I can do is refer to a sound young kids sometimes make when imitating someone who's supposed to be really, really stupid. Often they'll say *duh* or *guh*, and the first sound will be an implosive stop (so not regular *duh*). If this isn't bringing anything to mind from your childhood, bravo! But you'll have to seek elsewhere for a good example of how this is pronounced. If you live in Southern California, Vietnamese has two high frequency implosives that you can hear: *b* [ɓ] and *đ* [ɗ]. Growing up, my friends and I learned a *very* bad swear word beginning with [ɗ] that we would frequently try to imitate—ironically enough, producing [t’] more often than [ɗ]. (No, I won't say what it was; yes, I'll wait while you google it.) Otherwise, YouTube is your friend. I won't lie: implosives are tough. It takes practice to fit them in between vowels. It's best to hear a fluent speaker and try your best to imitate them.

Now for the sounds *everyone's* always asking about: Clicks. First, yes, there are languages that use clicks as ordinary consonants. Second, yes, most of them are in southern Africa. That said, most languages on Earth feature a click or two of some kind, just not as a consonant. Instead, they have affective uses (used to imitate other noises; used with animals; used to convey some sort of emotion, etc.). Just because they're rare, though, doesn't mean you can't use them.

A click consonant is one that is produced in a manner similar to an ejective. Remember that with an ejective consonant, airflow is closed off at the glottis, trapping air between the glottis and wherever the secondary closure is at. A click consonant is similar, except that air is closed off at the velum or uvula. Air is then trapped in between that closure and the secondary closure, which is always anterior to the velum. Before the anterior closure is released, air is sucked in, the result being negative pressure inside the little pocket made inside the tongue. When the anterior closure is released, air rushes in, which produces the clicking sound. Think of the sound we make when imitating a ticking clock (for younger readers, clocks used to tick. Ask an adult). Those are clicks. Try slowing down the pronunciation of one of those ticks and you should be able to track the motion described above fairly easily.

The IPA symbols that exist for clicks are pretty strange. Here are the basic click symbols:

Bilabial	**Dental**	**Alveolar**	**Alveolar Lateral**	**Retroflex**	**Palatal**
ʘ	ǀ	ǃ	ǁ	‼	ǂ

Unfortunately these don't tell the whole story. For just like stops, clicks can appear with a whole host of secondary features, including: voicing, palatalization, labialization, glottalization, epiglottalization, breathiness, nasality… They can even occur in pairs with other consonants, producing affricate-like contour clicks. While a number of languages with clicks have just a few, several of them have literally dozens—and most all of them feature different series (i.e. not just a row of clicks, but a

corresponding set of voiceless and voiced clicks, or voiceless, nasalized and aspirated clicks, etc.).

In trying to pronounce them, forget all about the different types of phonation, aspiration, etc. Instead focus on the place of articulation and try to produce the click in whatever way you can. Starting with the bilabial click [ʘ], this is similar to a kissing noise, except you don't bring the sides of your lips together (think of it as making a kissing sound while using the same mouth shape you would for a [b]). The sound should be a little noisy, a little wet, a little fricative-y—but it should sound markedly different from either a [b] or a [β].

I've already described the tick tock [ǃ] sound, which should sound much hollower and less noisy and fricative-y than [ʘ], so now try [ǀ]. Start by placing the tip of your tongue *right* behind your upper teeth. Make a closure, build up the pressure, then let it go. This sound should be halfway in between [ʘ] and [ǃ]. It should have the wet, smacking sound of [ʘ], but the alveolar-like sound of [ǃ]. This is a sound that some will use to, for example, reprove a cat. Myself, I actually do [ǃǃ] for that sound. For yourself, imagine a kitten has gotten after an entire roll of paper towels, and there are towels all over the floor. Now the kitten is cute, of course, so it's adorable, but at the same time the little raggamuffin did something wrong, so you go *ntch, ntch, ntch!* at it. That sound is either [ǀ] or [ǃǃ], depending on how you do it.

The alveolar lateral click [ǁ] is most commonly associated with the sound cowboys in the movies make when they're about to say *Giddyap!* to their horse. Though few of us are cowboys (and fewer still movie cowboys), I imagine all of us have experience making this click. Now you know how to write it.

The last click, the palatal click [ǂ], doesn't really have a place in English. Somehow it brings up associations of someone giving someone else who's good looking a little *hotchacha!* type of thing while making this noise, but I'm not sure how commonplace that is. To produce the click, place the tip of your tongue *firmly* behind your lower teeth (this isn't necessary to produce the noise, but it'll make sure your tongue tip plays no part in the click). Once you've done that, do the best click you can without using your lips. If you've followed the instructions exactly, you should be producing the palatal click [ǂ].

Now before attempting to try voiced, nasal and glottalized clicks, just try to put it in a word. Start with something simple, like putting it in between [ɑ] vowels (like saying *aaaaahhhh* at the doctor's), as shown below:

[ɑʘɑ]

[ɑlɑ]

[ɑ!ɑ]

It might seem at first like you're producing three separate unconnected sounds, but keep at it, and you should be able to pronounce and think of these as words, just as you would [ɑdɑ], [ɑnɑ] or [ɑpɑ].

Once you have that down, try adding other elements to the click. For me, nasalized clicks are the easiest type of click to do. Focus on allowing air to pass through your nose during the click while pronouncing the words as usual. Then try voicing and devoicing them; aspirating them; adding lip rounding, etc. The transcription for different types of clicks can get pretty tricky, and there are varying standards, so a good description is always advised if you plan to use a large number of clicks in a conlang.

Chances are the conventions outlined below will be outdated twenty years from now, but for the interim, here are different types of clicks (using the bilabial click [ʘ] as a base):

[ʘ] = basic (or tenuis) click

[ʘʰ] = aspirated click

[ᵍʘ] = voiced click

[ᵑʘ] = nasalized click

[ʘˀ] = glottalized click

[ʘ͡q] = contour click (click symbol followed by uvular sound with tie bar)

Older click terminology would put a velar consonant before the click which would describe the phonation, etc. (e.g. [kʘ], [gʘ] [ŋʘ]…). This practice has been deprecated, but you may run into it online.

If you plan to try to use clicks in a conlang, I *strongly* encourage you to listen to them used in fluent speech. To do so, find pretty much any language on Wikipedia that features click (focus your search on southern Africa), then type the name of the language into YouTube. I've found some wonderful examples that way. Looking at the transcriptions of languages far different from your own, though, can kind of turn your head to soup. It's like looking at calculus equations. Hearing the actual sounds in fluent speech really helps to ground it—and that goes double for clicks. It may seem very difficult, but it's doable, and hearing someone speaking a language with clicks in it fluidly is really quite beautiful.

Secondary Articulations

Now that we've discussed the major consonant groups, I'd like to briefly touch on secondary articulations. Secondary articulations are various types of modifications that aren't full consonantal features in and of themselves, but which can be used to produce meaning distinctions—or dialectal variation—in other consonants. One we've already seen that should help clarify what I mean by secondary articulation is aspiration.

Aspiration occurs when the glottis is left open so that air can pass through during the production of another sound. In English, voiceless stops are aspirated at the beginning of a word, and in a few other places. In the introduction, I gave you a test to illustrate the difference between *top* and *stop*. Now I'd like to follow up and tell you a bit more about aspiration.

First, pretty much *all* sounds can be aspirated. When aspiration doesn't play an important role in distinguishing meaning, it's typically not written (this is the case with English, German, Japanese and other languages where voiceless stops have some amount of aspiration). In certain languages it's quite important, though. I showed you one example from Hindi, but here's another example:

पाट [paːʈ] "span"

फाट [pʰaːʈ] "division of land"

बाट [baːʈ] "dumb bell (weight)"

भाट [bʰaːʈ] "bard"

As you can see, the only difference between these four is voicing and aspiration. Using aspiration as a distinguishing feature is fairly common, and a conlanger should

feel comfortable employing it. Having said that, the phonetic details are usually a little muddier. For example, while the four way stop contrast is fairly strong phonologically, phonetically the aspirated voiced sounds are probably better described as murmured rather than truly aspirated. Even so, it's still considered acceptable to write the voiced stops as aspirated in Hindi unless you're doing a detailed phonetic study.

To write an aspirated sound, you simply add a little [ʰ] right after the sound. You will sometimes see a little voiced [ʱ] used with voiced sounds, but it's not as common, even if it is more accurate phonetically. Incidentally, you should try aspirating sounds you never would have thought of aspirating before. For example, [sʰ] is found in Korean, and is really fun to pronounce.

I'll now give you a list of common secondary articulations (plus a couple other consonant modifications) and how to write them. Below, [m] will just be used as a placeholder. A secondary articulation will usually be able to accompany any consonant:

[mʰ] = aspiration

[mʱ] = voiced aspiration

[ʰm] = preaspiration (aspiration before the consonant is pronounced)

[m̄] = no audible release

[mː] = length marker (sometimes used to mark geminates)

[mʲ] = palatalized (e.g. some pronunciations of *new* [nʲuː]

[mʷ] = labialized (lips are rounded when this sound is pronounced)

[m̤] = breathy voiced (like Hindi voiced aspirated stops)

[ⁿm] = prenasalized (cf. Swahili *Ndio* [ⁿdio] "yes")

[mⁿ] = nasal release (like the *dn* in *madness*, but much quicker)

[mˀ] = glottalization (often identical to an ejective)

[m’] = ejective

[mˤ] = pharyngealization (back of the tongue touches the pharynx)

[mˠ] = velarization (back of the tongue touches the velum)

[mˡ] = lateral release (think the *tl* in *chipotle* or *Nahuatl*)

[m̩] = syllabic (consonant serves as a vowel, as in *button* [bʌʔn̩])

[m̥] = voiceless

[m̰] = creaky voice (popularly called vocal fry)

[m̃] = nasalized (most often applied to vowels)

[m˞] = rhotic (sound is somehow more *r*-like than usual)

There are actually dozens more than this, but these are the most commonly-used IPA diacritics. Keep all this in mind as we move on to vowels!

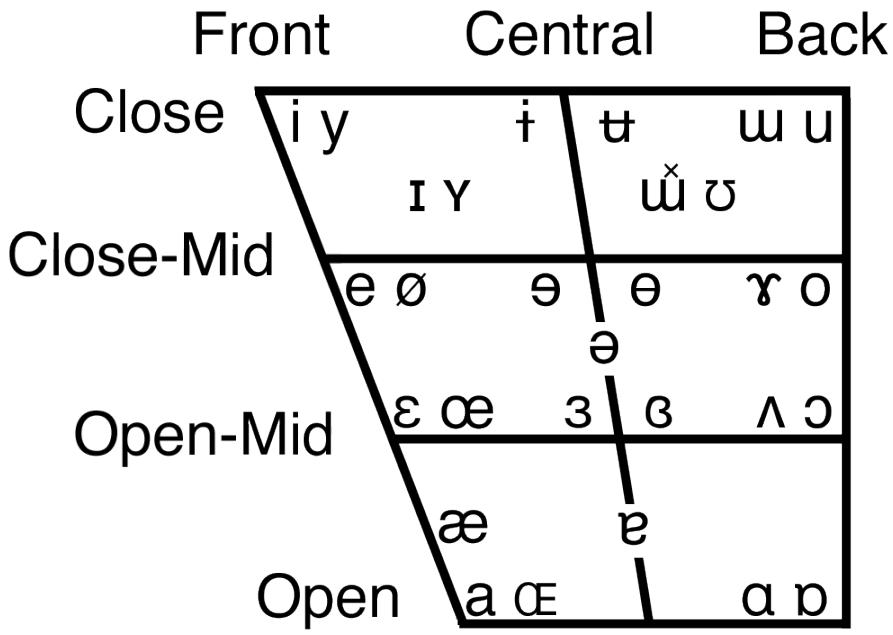
Vowels

Wherever you are (especially if you're in a library, bookstore or park), open your mouth and scream. *Loud. SCREAM!!!*

That's a vowel.

A **vowel** sound is produced when air is allowed to pass out of the lungs totally unimpeded. We can move our tongues and lips around and even lower our velums and wiggle our epilogottises to alter the sound of the vowel, but so long as air is allowed to pass out of the lungs in an unimpeded stream, the resulting sound is a vowel. Vowels are among the most fluid sounds we have. They're *desperately* hard to pin down. This is why the *i* in *burrito* is pronounced differently in English and Spanish, *even though they're the exact same sound*. It's easy to move your tongue *just* a little bit and change the overall character of a vowel. It's also the hardest thing to approximate. Those who are able to do other accents well or who can make themselves sound like a native when speaking another language are incredible at imitating other vowel sounds. Consonants are a cake walk with no admission fee compared to vowels.

Unlike consonants, which need to be dealt with based on their manner of articulation, we're going to go ahead and look at all the vowels at once. Below is a chart of all the major vowels found in the world's languages:



The first question you'll probably have after seeing that chart is why is there an oddly-shaped quadrilateral behind the vowel symbols. The reason has to do with how vowels are described by linguists. All vowels are defined by three basic measures: **backness** (whether the tongue body is closer to or further away from the pharynx); **height** (whether the tongue body is close to the roof to the roof of the mouth or close to the bottom of the mouth); and **rounding** (whether the lips are rounded or not when pronouncing a vowel). The symbols are placed where they are to mirror the position of the tongue body. Consequently, the tongue body is quite high and forward in pronouncing a vowel like [i], but quite low and back when pronouncing a vowel like [ɑ]. You may also notice that your mouth is far more open for the vowel [ɑ] than it is for [i]. **Openness** is another way of describing vowel height. In other words, rather than noting that the tongue body is low in the mouth for [ɑ], you'd say the jaw is open. For our purposes, these two systems shall be treated as identical.

You can actually hear the effect of height and backness by doing two clever tricks that my phonetics professor John Ohala taught me. To hear the effect of height, what you should do is mouth, in order, the vowels of *meet* [i]*, mate* [e]*, met* [ɛ] and *Matt* [æ]. Don't actually pronounce them. While doing this, take your thumb and forefinger and flick the skin underneath your back jaw. This will produce a hollow popping sound (something like clapping your palm over an open bottle). As you move from the higher vowel to the lower vowel, the tone of that hollow popping sound will actually get higher. You can repeat the example with *moot* [u], *moat* [o] and *mot* [mɑt] and hear the same result.

For backness, you can actually whisper the vowels. When whispering (no vocal fold vibration at all), you can't affect the tone of your speech at all. The inherent tone of each vowel *decreases* the further back in the mouth the tongue body goes. So if you whisper the vowels for *meet, mate, met, Matt mot, moat* and *moot* in that order, the pitch will get progressively lower. Neat, huh?

The last feature is lip rounding, which doesn't really need a trick, as it's pretty easy to tell when you're lips are rounded and when they aren't. You should notice that your lips round quite a bit when pronouncing *moo* [muː], but unround completely when pronouncing *me* [miː]. That's fairly standard. Now try to do the opposite. Keep your lips *completely* unrounded and pronounce *moo*. If you succeed, you should be pronouncing [mɯː] (and sounding like a stereotypical Southern Californian in the process). Now try *me* with your lips *completely* rounded. If you succeed, you'll be pronouncing [myː]. It may sound like a strange sound, but it's found in French, German and Turkish—and also English, in certain circumstances.

You know how at sporting events it's not uncommon to chant the name of the home team in one fashion or another? Think about the teams you know that have a main [i] vowel. One that always comes to mind for me is the Miami Heat. When fans chant "Let's go Heat!", more often than not the name of the team comes out [hyt] rather than [hit]. The reason is that the lip rounding makes the sound more resonant: it lengthens the tube that is our vocal tract, so to speak, and allows the speaker to give more volume to what they're saying. This is also why players with [u] in their name are much more likely to have their name chanted.

With this information in hand, you should be able to pronounce *any* vowel. It also should become clear how these vowel symbols are really just buoys on the vast ocean of vocalic possibility. I bet right now without even trying you can produce five or six different sounds that would qualify as [i] but which are all slightly different. The way a specific language works is there's an entire range of sounds that qualify as a particular sound, and so long as you're somewhere in that range, you're fine. If you cross the boundary, though, you'll either be pronouncing a different vowel or a strange sound that will ring false in the ear of a native speaker.

First, since English has a goodly number of vowels, let's go over them so you can note the different and use these as reference. I'm assuming a Standard American pronunciation (so not even my pronunciation). If you need a reference for some of these words, think of a stuffy male news anchor pronouncing these words, and that should be a good enough frame of reference:

Front Vowels	Back Vowels
bead [bid]	*booed* [bud]

bid [bɪd] *hood* [hʊd]

bade or *bayed* [bed] *bode* [bod]

bed [bɛd] *baud* [bɔd]

bad [bæd] *bod* [bɑd]

English has a couple other vowels worth noting. The first vowel in *bud* and the second vowel in *sofa* sound quite similar to me, but linguists insist on transcribing these words [bʌd] and [sofə], respectively. I'll leave you to decide. The sounds [ʌ] and [ə] are really quite different, so I don't see it, but I'm doing my duty and passing on the convention that the vowel in question is transcribed [ʌ] when stressed and [ə] when it isn't. In addition to this vowel, English also has [ɚ], which we've seen before. That's the vowel in *bird* [bɚd]. (Notice that linguists don't make a big noise about transcribing that [bʌ˞d]. Hypocrisy, I say!)

With that inventory in mind, all the other vowels can basically be described in reference to English. For example, [y], [ʏ], [ø] and [œ] are simply the vowels [i], [ɪ], [e] and [ɛ] pronounced with fully rounded lips. The vowel [a] is like [æ], but lower (try opening your mouth wider—like a stereotypical Texan pronunciation of a word like *light*). The vowels [ɯ], [ɯ̹], [ɤ], [ʌ] and [ɒ], on the other hand, are just like [u], [ʊ], [o], [ɔ] and [ɑ] pronounced with fully unrounded lips.

The central vowels may take a little work, since we've just got [ə] in English. Think of how that vowel is pronounced, though. Channel the feel of that vowel. Now without moving your tongue forwards or backwards, try moving your tongue up and down. Add or subtract lip rounding as necessary, and that should give you vowels like [ɐ], [ɨ], [ʉ], [ɜ], etc. That last vowel is what's typically used for the standard English

pronunciation of *bird* [bɜd]. Try it out. Otherwise in English the place where you hear central vowels the most is in popular songs. This is a phenomenon that really came to promise in the 90s and appears to be here to stay.

Do you know the song "Better Man" by Pearl Jam? (No judgement if not; they've done better.) See if you can find it on YouTube. Listen to the part of the chorus where Eddie Vedder sings "Can't find a better man". Hear how his voices changes—how it kind of sounds huskier? This is something you heard a lot in the 90s (Scott Weiland did it; Shakira does it a lot; Dave Matthews did a lot [or Dave, as his true fans call him]). What Eddie Vedder is actually doing is centralizing all the front vowels. His typical pronunciation of "can't find a better" is something I'd transcribe as [kʰəːmfɐːnəbɜrə]. Naturally, he doesn't *always* sing this way. Every so often he simply feels a need to kick it into overdrive, and so centralizes all the vowels. It's a noticeably different sound. As for why, the only thing I can come up with is that it obscures a lot of the vocalic variety of English (there are fewer distinctions for central vowels than for front vowels), and makes it easier to hold a tone. It's also why *baby* comes out *babay* a lot of times ([e] is lower than [i], which means your mouth is open wider). Anyway, if you're trying to nail central vowels, remember Eddie Vedder (but hopefully for "Corduroy", "Yellow Ledbetter", "Black", "Guaranteed", "Oceans" and "I Got Id" rather than "Better Man").

Now that we have all the vowels down, I'd like to discuss some properties of vowels that certain languages will make use of beyond vowel quality. The first is **vowel length**. Compare the vowel in the English word *bat* to the vowel in the English word *bad*. Yes, the stops at the end will be different, but pay attention to the length of the vowel. Notice how the *a* vowel of *bad* takes more time to pronounce than the *a* sound in

bat. This is because vowels are naturally lengthened before voiced consonants in English—and, in narrow vocalic transcription, we'd transcribe those two words [bæt] and [bæːd], respectively. You can try this experiment with any vowel pair in English, and the distinction should hold true: *bit*/*bid*, *rot*/*rod*, *neat*/*need*, *loot*/*lewd*, *mate*/*made*, etc. The mark we use to indicate a long vowel is a kind of modified colon that looks like this: [ː]. You place that mark after a vowel to indicate that it's long. You can also use one little triangle after a vowel [ˑ] to indicate that a vowel is *slightly* longer than a short vowel.

Unlike English, though, which uses vowel length phonetically, many languages use vowel length to distinguish meaning. Among natural languages, Hawaiian, Arabic, Japanese, Hungarian, Finnish and Latin distinguish long and short vowels. Amongst conlangs created for television and film, High Valyrian from HBO's *Game of Thrones*, Shiväisith from Marvel's *Thor: The Dark World* and Lishepus from Syfy's *Dominion* contrast long and short vowels. Here's an example from High Valyrian:

kelin [kelin] "I stop"

kēlin [keːlin] "herd of cats"

Estonian is a language that's famous for making a three-way length distinction, even though the distinction isn't reified in the language's official orthography. Estonian distinguishes short vowels, half-long vowels and long vowels, albeit the jury's still out on whether or not there's something else at work in the so-called "overlong" vowels. Here's an example:

sada [sata] "hundred"

saada [saˑta] "send (command form)"

saada [saːta] "to get"

As you can see, the correct pronunciation is simply something you'll have to learn to speak fluent Estonian, as the half-long and long (or long and overlong) vowels are written the same way. Consistent distinctions beyond a three-way length distinction are unheard of in natural languages, but length is often used for other purposes (e.g. [ʏːːːːːːːːː] which is how I'd transcribe myself saying, "*Ohhhhhhhhhh…*").

Another quite common use of vowels is the **diphthong**. A diphthong is a vowel which starts out as one vowel but finishes as another. Consequently, it doesn't sound the same going forwards and backwards. For example, trying slowing down your pronunciation of the word *lie*. Notice where your tongue is when you start the vowel of that word and where it ends up. At the beginning, you should be pronouncing a vowel very much like [a], and at the end it should sound a lot like [i]. Despite that fact, we treat this as a single sound (English speakers will think of this as a "long *i*"). That's what a diphthong is: a dynamic vowel that is treated as a single vowel. So, for example, the *i* in *light* [lajt] is a diphthong, as is the *ow* in *how* [haw], but the *ea* in *react* [riækt] is just two vowels coming one next to the other.

There's some controversy about how exactly to transcribe diphthongs. Using the *i* in *light* diphthong as an example, it's sometimes transcribed as a vowel followed by a glide: [aj]. The idea behind this is if you want to emphasize the vowel (imagine pronouncing the word very slowly for someone who's having trouble understanding), you'll lengthen the [a] part, and the off-glide will be short. Consequently it's treated as a consonant. I generally favor this method of transcription, but it can be confusing, as the thing is a vowel, not a vowel followed by a consonant.

Others have favored the dual vowel approach, writing either two full vowels (i.e. [ai]), or a vowel followed by a lax vowel (i.e. [aɪ]). I've never been a fan of this method, and so avoid it if possible. Nevertheless, you will see it, so you should be aware of it.

However you describe it, it's important to note in a conlang's description what the diphthongs are, how they're pronounced, and how they're going to be transcribed.

Also, as a note, vowels are, by default, voiced. Certain languages, though, have made use of voiceless (or whispered) vowels. For example, in Japanese, the high vowels [i] and [ɯ] tend to be voiceless in between voiceless consonants. This is why names like *Daisuke* appear to be pronounced [daiske]. In fact, a name like that will be pronounced [daisɯ̥ke], with the vowel being whispered in between the two consonants. A similar thing may happen in certain pronunciations of English words like *pecan*, if the stress is on the second syllable. The first vowel won't actually be voiced, getting swallowed up by the voicelessness surrounding it, resulting in a pronunciation like [pʰə̥kʰɑn]. Despite their presence in natural language, no language yet features a full set of voiced and voiceless vowels that can appear in any context. Such a thing would probably never occur in a natural language, but could happen in a conlang.

Other terminology you'll see with respect to vowels are terms like **tense** and **lax** as well as the feature [ATR], which stands for "advanced tongue root". These are generally properties of sound systems rather than vowels themselves. So, for example, in English, [i], [e], [u] and [o] are considered tense vowels, whereas [ɪ], [ɛ], [ʊ] and [ɔ] are considered their lax counterparts. Not all languages will divide those up the same, so when you see features like "tense" and "lax" being used, be sure to investigate the

phonology of the language, as what's tense in one will not necessarily be tense in another, and the same goes for laxness.

Finally, a common feature of vowel systems is nasality. We've already discussed nasal consonants, so you understand how nasality works. Now apply those same principles to vowels. Try pronouncing a nice [ɑ] vowel with your velum lowered. This will mean that air will primarily be passing out of your lungs and out through your nose. If you do this successfully, you'll be pronouncing [ɑ̃], which is precisely how you pronounce French *an* (one of the words for "year" in French). Nasalization can be applied to any vowel, but most languages that employ nasal vowels only allow a subset of nasal vowels. French, for example, has eleven oral vowels (or regular vowels), but only three nasal vowels. This is usual, but not necessary. It's perfectly possible to pronounce any vowel as oral or nasal: just pronounce the vowel with a lowered velum. (See the Gbe languages for an example of a language that has the same number of oral and nasal vowels.)

It takes practice to be able to produce, distinguish and remember the various sounds we've discussed thus far. Please feel free to use this section as a reference guide that you can refer to in future sections, as we'll be relying on your understanding of the phonetic principles of language as we move through the book. This is just the foundation you need to have down so we can get to the good stuff!

Phonology

So far we've been discussing phonetics and phonetic detail. Once you hit the ground running with actual language, though, you'll discover that language users work with language at some level that is *less* precise than what we've seen thus far. For example, take the English word *negative*. Ordinarily we distinguish the vowel in *met* [mɛt] and the vowel in *mate* [met]. Those are two different words. Now try pronouncing *negative* with both vowels—i.e. as [negəɹɪv] and [nɛgəɹɪv]. You should notice a difference, but is it an important one? Probably not. You might not even notice it if someone pronounced one over the other. In fact, I'd go so far as to say there could *never* exist an English word that would make such a distinction before [g]. Isn't that weird? We have *mech* [mɛk] and *make* [mek], but try distinguishing *Meg* [mɛg] from *mague* [meg] and see how far you get.

And yet I'd like you to try pronouncing *vague* [veg] as [vɛg]. *How does that not work?!*

What you're noticing here is the effect of **phonology**. The phonology of a language is an abstract layer of understanding which treats actual sounds (**phones**) as subsets of other sounds (**phonemes**). The phonology of English is the reason we think of the *t* in *stop* as identical to the *t* in *top*, even though the former is [t] and the latter [tʰ], and there will be languages that treat them differently. We consider both sounds to be realizations of the phoneme /t/ (recall that phonemic transcription is written between forwards slashes).

Throughout the rest of this sections we'll be discussing phonology and phonological phenomena. This is where things get interesting.

Sound Systems

You've seen the myriad speech sounds available to the human mouth. Being the speaker of a language (it doesn't matter which), you also should have noted that your language doesn't utilize *every possible* speech sound. This is true of every language on the planet. How a language chooses sounds is a bit of a mystery (the history is lost to antiquity), but as a conlanger, it's on you to choose the sounds for your language. The result will be your conlang's **sound system**. If you're creating a naturalistic language, there are a number of principles that will help to guide you in doing so, and this section will detail those principles.

First, take a look at the phonology of *any* language. For example, let's look at the consonantal inventory of Tukang Besi, an Austronesian language spoken in Indonesia whose reference grammar was written by superlinguist and martial arts expert Mark Donohue (he is literally both of those things).

Manner	**Bilabial**	**Dental/ Alveolar**	**Velar**	**Glottal**
Oral Stop	p, b*, ᵐp, ᵐb	t, d*, ⁿt, ⁿd	k, g, ᵑk, ᵑg	ʔ
Implosive	ɓ	ɗ		
Nasal Stop	m	n	ŋ	
Fricative	β	s, ⁿs, z*		h
Trill		r		
Lateral		l		

Note that sounds marked with an asterisk only appear in loanwords, and everything other than [r], [s] and [n] in the **Dental/Alveolar** column is dental. So what can we say about this? First, it's worth noting that [b] and [d] aren't native, and that there

are only two implosives: [ɓ] and [ɗ]. And the one place where there isn't an implosive (the **Velar** column), there is a native [g]. Other than that, notice how balanced everything looks. There are basically four native stops in each major place of articulation: voiceless, voiced (either implosive or regular), and both prenasalized and plain. There's one *r* sound, one *l* sound, a couple glottals, and then a sibilant (or strident) fricative [s] and a weaker voiced one [β]. This looks natural. This, however, would strike me as bizarre:

Manner	**Bilabial**	**Dental/ Alveolar**	**Velar**	**Glottal**
Oral Stop	^{m}p	t, d, ^{n}t, ^{n}d	k, g, ^{ŋ}k, ^{ŋ}g	ʔ
Implosive		ɗ		
Nasal Stop	m		ŋ	
Fricative	β	s, ^{n}s, z	x, ɣ, x^{w}, ɣ^{w}, x^{j}, ɣ^{j}	h, ɦ
Trill		r		
Lateral			ɬ̥	

This is basically the same phonological inventory with a couple changes. Notice that there's now exactly one bilabial stop, and it's a prenasalized voiceless bilabial stop [^{m}p]. There's also only one implosive, and it's dental. There is no dental or alveolar [n], for apparently no reason; there's a full series of plain, labialized *and* palatalized velar fricatives (there are no other labialized or palatalized consonants); and there's only one *l*-sound, and it's a *voiceless* velar lateral.

This should strike you as exceedingly unnatural. That is, even though all of these are sounds human beings can produce (and without too much difficulty, I might add), I would *never* expect a language to exist that had precisely this phonological inventory.

The reason behind the non-occurrence of sound systems like this one is a principle that I call **acoustic economy**.

Acoustic economy is, simply, the idea that languages will conspire to take maximal advantage of the sounds available to human beings. They will do so *not* because certain sounds are more difficult to *pronounce* than others, but because certain distinctions are more difficult to *hear* than others. Let me illustrate what I mean by this.

In English, we distinguish *t* from *k* from *p* well enough (e.g. *kick* vs. *tick* vs. *pick*). Consider the word *September*, though. Would you expect there to exist an entirely separate word that was spelled *Sektember* (or maybe *Sectember*. Yeah, that looks more Englishy)? This would be a word totally unrelated to *September*, mind. This is certainly something English *could* do, of course, but the place where the *p* occurs is, for a plethora of reasons, a really poor environment to distinguish those otherwise easily distinguished sounds. Again, it's easy enough to pronounce *Sectember* (in fact, that *is* how my late stepfather used to pronounce *September*), but you have to think about it from the listener's perspective. In noisy environments, with a word like this pronounced lazily on occasion, having to distinguish *September* and *Sectember* faithfully wouldn't work.

Now if you can pull back and imagine a language evolving over thousands of years, and there being billions of word pairs like this, and billions of interactions amongst billions of different speakers, you should be able to begin to understand the role that acoustics plays both in the organization and reorganization of sound systems, and in historical sound change (which we'll discuss in detail later on). A user of a language reproduces the language in the way that they believe it's supposed to be

used, and that's based on what they hear—and also what they read, if the language has a written form. Consequently if a certain distinction is routinely difficult to perceive, it can eventually collapse.

Of course, something that will rescue certain sounds that might seem difficult to perceive in certain environments is having a system behind them. Think of this as safety in numbers. So, for example, let's take a look at Hindi again. Here are the stops and affricates of Hindi:

Voicing	**Bilabial**	**Dental**	**Retroflex**	**Palatal**	**Velar**
Voiceless	p, pʰ	t, tʰ	ʈ, ʈʰ	tʃ, tʃʰ	k, kʰ
Voiced	b, bʰ	d, dʰ	ɖ, ɖʰ	dʒ, dʒʰ	g, gʰ

From the point of view of a Hawaiian speaker, this looks like a nightmare. Hawaiian doesn't even distinguish voiced and voiceless stops, let alone aspiration! But given that this is obviously a *system* with voicing and aspiration distinguished at every place of articulation where there's a stop, it makes sense. Imagine if instead the inventory of Hindi looked like this:

Voicing	**Bilabial**	**Dental**	**Retroflex**	**Palatal**	**Velar**
Voiceless		t	ʈ	tʃ, tʃʰ	k
Voiced	b	d	ɖ	dʒ, dʒʰ	

Outside the palatals, that's actually not an unrealistic system. The fact that palatals distinguish voicing *and* aspiration, and that this is the only point of articulation where this happens, though, makes this system look completely unsustainable. Again, it's possible to produce these distinctions, but it seems highly unlikely that it would remain this way for any appreciable length of time.

Of course, there are always exceptions to the rule. Take a look at the stops of Vietnamese:

Voicing	Bilabial	Dental/ Alveolar	Palatal	Velar
Plain	p	t	c	k
Implosive	ɓ	ɗ		
Aspirated		tʰ		

Yes, that aspirated stop sticks out like a sore thumb (or a sore tʰumb [I apologize for nothing]). And yet, there it is. Even so, there are several mitigating factors to consider. For one, both [t] and [ɗ] are alveolar, while [tʰ] is actually dental. It's the only stop that's dental, but not the only dental sound in the phonology, so it is a part of a series. Second, it's important that this sound is **coronal**. A coronal sound is a sound with a dental, alveolar or post-alveolar place of articulation. There tend to be more distinctions made crosslinguistically for coronal sounds than there are for any other place of articulation. It would be *much* more bizarre if the only aspirated sound was bilabial. And furthermore, the aspiration helps to distinguish the dental [t] from the alveolar [t]—a distinction that otherwise is difficult to parse consistently. Consequently, maximizing the difference between the two stops helps to preserve them as separate phonemes. Had there been no other distinguishing factors, I imagine the distinction would have collapsed.

The takeaway from a sound system like that of Vietnamese, though, is there tends to be an explanation for features that appear to be out of the ordinary. That is, while it's odd that Vietnamese has a dental aspirated stop, it doesn't boggle the mind

the way a voiceless bilabial stop would. Often the explanation is that the sound has been borrowed. For example, in English a word can't end with a stressed [ʌ] or [ə] (i.e. the *u* vowel in *bud*). Now, though, a relatively newly borrowed word, *pho*, is being pronounced closer to its original pronunciation as [fʌ], the older pronunciation [fɑ] or [fɔ] now falling out of favor (at least in Southern California). A word like [fʌ] should be impossible in English. The only reason it exists is because we borrowed the word from Vietnamese, and are actually doing our best to pronounce it somewhat accurately. Thus a new pattern is introduced to the language.

The principle of acoustic economy applies to vowels as well as consonants. Here's a look at the vowel inventory of Hill Mari, spoken in Russia:

Height	**Front**		**Central**	**Back**	
	Unround	**Round**		**Unround**	**Round**
High	i	y			u
Mid	e	ø	ə	ɤ	o
Low	æ			ɑ	

This is a diverse system, but not at all out of the ordinary. Notice that there's a lot of diversity amongst the mid vowels, but less amongst the high and low vowels. In general, it's standard for front vowels to be unrounded, and for back non-low vowels to be rounded, so while the system is certainly a bit odd to an English speaker, it perfectly natural.

Here's a system that should look pretty odd:

Height	**Front**		**Central**	**Back**	
	Unround	**Round**		**Unround**	**Round**

High		ʏ			
Mid		ø	ɵ	ʌ	
Low	æ	œ	ɐ	ɑ	ɒ

Here again we have a system that is certainly possible for a human being, but if one ever came across a language on Earth that had this exact vowel inventory, it would prove fairly devastating to phonological theory. There's only one high vowel, and it's a front rounded lax vowel; there's enormous diversity in the low vowels, including the low round [œ] vowel, which has not yet been found in any language; the only two mid vowels are opposite rounding vowels… It's a mess.

As with consonants, the principle of acoustic economy expects for languages to maximize the phonological space available, so that words are audibly distinct. If they're difficult to distinguish, the distinctions tend to collapse, ironed out by evolution.

Competing with the idea of acoustic economy is something I call the principle of **brand identity**. In marketing, the goal is to make sure that every piece of information related to a brand has the characteristics of its brand identity on it (logos, color schemes, slogans, font faces, etc.). You should be able to look at anything associated with a particular brand and tell it's from that brand. In language, the same principle applies, albeit a little differently.

Looking at sound systems, there are certain phonemes that have a rarer distribution crosslinguistically than others. For example, it's much more common to have a /t/ than a /ʔ/ in a language. When there is a sound that's out of the ordinary, though, it usually brings a long a whole host of brothers and sisters with it. Arabic, for example, has a pharyngeal approximant [ʕ]. It's not a common sound, but not vanishingly rare. To

bolster it, though, it also has a voiceless pharyngeal fricative [ħ], two pharyngealized alveolar stops [tˤ] and [dˤ], and two pharyngealized fricatives [sˤ] and [ðˤ] (or, in some dialects, [zˤ]). If you have trouble pronouncing a pharyngeal approximant, you're in for a fun time if you want to learn Arabic!

Looking at an example we've seen before, retroflex consonants are fairly rare crosslinguistically, but are present in Hindi. Remembering that Hindi features distinctive aspiration, here is the full set of retroflex sounds found in Hindi: [ʈ, ʈʰ, ɖ, ɖʰ, ʂ, ɽ, ɽʰ, ɳ]. In addition, Hindi has a series of nasal vowels: [ɑ̃, õ, ɔ̃, ũ, ĩ, ẽ, ɛ̃].

This is a phenomenon you'll see in language after language. It may be difficult to distinguish [t] from [tˤ] or [o] from [õ], for example, but if the language is going to do it, it may as well do it a *lot*. This actually helps to preserve the distinction. That is, if there are a *class* of pharyngealized sounds, or palatalized sounds, or glottalized sound, or what have you, speakers and listeners get constant interaction with the phenomenon. If one is able to distinguish a nasalized vowel from an oral vowel, then it's not a problem to distinguish a *particular* nasalized vowel from its oral counterpart.

This is what I mean by branding identity. A language will take advantage of its unique sounds and make them a hallmark of the language. From English, consider our rarer sounds: [θ, ð, ɹ]. The sound [ð] isn't in a lot of words, but it does find itself in a lot of *really* high frequency words: *the* [ðə], *this* [ðɪs], *that* [ðæt], *then* [ðɛn], *though* [ðow], etc. The sound [θ] is common enough, and is used in the -*th* ending in words like *width*, *warmth*, *depth* and in the ending for ordinal numbers like *fifth*, *eighth*, *twentieth*, etc. And you can't get through a sentence without using an *r*. It's everywhere!

In building the sound system for a naturalistic language, then, keep these two principles in mind. As you build, though, one question might emerge that I can address—namely, how big or small does one's sound system have to be?

On the high end, my suggestion is not to worry about the number of consonants or vowels. Instead, the question should be: Does my system make sense? For example, the !Xóõ language has over a hundred consonants (the majority of them clicks), but the distribution of consonants is still quite principled. The same is true of all natural languages.

Now, if you go the *other* direction, that question is quite interesting.

The natural language with the fewest number of consonants is a language of Papua New Guinea called Rotokas. Depending on the dialect and the analysis, Rotokas is analyzed as having either six or nine consonants—and that's it. They're basically /p, t, k, b, d, g/ (don't let the spelling of the name of the language fool you. They use the Roman alphabet and have some funky spelling rules). It has ten vowels (five qualities, long and short), so there's plenty of syllabic possibilities (which is probably the real question that needs to be answered), but it looks like the answer to the question of how many consonants does a language *have* to have is six. You might be able to go smaller, but doing so would raise an eyebrow or two.

As for vowels, there are a group of languages all found in the Caucasus Mountains that are famous for having, depending on the analysis, exactly two vowels. Some argue there are three; some argue there are more than that. Nevertheless, a two vowel analysis has been made for a number of languages in this region—in particular, Ubykh and Kabardian, though Abkhaz and Adyghe are often thrown in. What are the

two vowels? It may surprise you: /a/ and /ə/. Each vowel ends up being realized in a lot of different ways depending on the consonants surrounding it, but since the realizations are consistent, the vowels are analyzed as being just /a/ and /ə/, while each of these languages has an *extremely* large consonant inventory. For example, Ubykh has an entire series of labialized consonants. When a labialized consonant precedes the vowel /a/, the result is the consonant followed by [o]. When it precedes /ə/, the result is the consonant followed by [u]. So, coming up with a nonce example, if you had the sequence /dʷə/, it would be pronounced [du]. Call it cheating if you will, but I've seen one of these analyses, and the logic is sound. I'm totally behind the two vowel analysis.

Now having said this, most languages have between four and six vowel qualities (not counting long vowels as separate), and between 20-30 consonants. English has an average number of consonants and an above average number of vowels. A language like Spanish is much more ordinary, with between 18 and 20 consonants, depending on the dialect, and five vowels. That should give you an idea of what the usual bounds are for natural languages. So long as you're aware of them, you can make a conscious decision to have your language fall within expected norms, or be an outlier.

In addition the foregoing, here are some quick tips if you're sitting down in front of a blank sheet of paper and creating your first sound system:

- Your system *will* be revised. Feel free to try things out to see how they work.
- Look at a variety of different languages' sound systems for inspiration, or just to get a better idea of what kind of variety there is. Wikipedia is great for this. Search for "[language name] phonology" (or if it's a lesser known language,

"[language name] language") and jump to the phonology section. Most pages have nice big charts to look at.

- Remember that sounds come in groups—especially **obstruents** (stops, fricatives and affricates). Don't go to add a palatal *sound*: go to add a palatal *series*.
- **Sonorants** (vowels, laterals, trills, flaps, approximants and nasals) tend to be voiced and tend not to have voiceless counterparts. Treat this as a default which can be circumvented if so desired.
- Bilabial obstruents are more likely to be voiced; velar/uvular obstruents are more likely to be voiceless. In between, distinguishing voicing is quite likely. This isn't a rule, but a tendency.
- Implosives grow increasingly less common starting from the lips and moving inwards; ejectives grow increasingly less common starting from the velum and moving outwards (i.e. outwards or inwards).
- **Sibilants** (*s*-like and *sh*-like sounds) are acoustically strong; non-sibilant fricatives are acoustically weak. It's not uncommon for a language to have one sibilant and one non-sibilant—and also not uncommon for non-sibilants to be confused for one another (e.g. [f] for [θ]; [h] for [ɸ], etc.).
- If a language has an opposite rounding vowel (i.e. front rounded or back unrounded), it will *almost* always have the regular version of the vowel as well. So if a language has [y], it will also have [i]. This isn't a rule, but a tendency.
- Low vowels, whether front or back, are most commonly unrounded. You will find [ɒ] in plenty of languages (it's the *o* in Alan Rickman's pronunciation of *Harry Potter*), but [ɑ] is much more common.

- *Always* sound things out. It will help you to understand the sound and its acoustic effect better.

To emphasize, just because you've selected the sounds you want to use in your language doesn't mean you're done with your phonology. Indeed, that's just the beginning!

Phonotactics

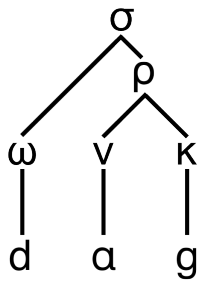
Take an English word like *strong*. It's a nice, sturdy English word. A phonetic transcription of it would be [stɹɑŋ]. The sounds that comprise that word are, without a doubt, English sounds. Consequently, one should be able to take those sounds and create a new word—say, a word like [tŋsɹɑ]. That makes a nice plausible English word, right?

Channeling a young Jennifer Connelly: Of *course* it doesn't!

The reason it doesn't is that the **phonotactics** of English simply do not permit a word like [tŋsɹɑ] to exist. If it had to exist, we'd probably end up pronouncing it [tʰʌŋzɹɑ]. In other words, we'd modify the word so that it obeyed the phonotactics of English. That's what the phonotactics are there for. They're a set of rules that tell the speaker of a language which types of combinations of sounds form coherent words, and which don't. This allows us to identify a language and ignore everything that doesn't sound like the language we're listening for.

Once you have your sound system, the next important step is deciding how the phonotactics of your language will work. In addition to the sounds present in the language, the phonotactic patterns present in your language are one of the three key factors in determining the phonaesthetic character of your language (this will be discussed in detail later). If phonotactics are ignored, more often than not the phonotactic patterns of a conlanger's native language are borrowed into their conlang, resulting in a conlang which behaves, for an English speaker, much more like English than the conlanger was intending.

The first step in creating a phonotactic system for a conlang is determining what constitutes a **syllable**. A syllable is a prosodic unit of measurement used to divide words into smaller parts. A syllable itself can be divided into two main parts: an **onset** and a **rhyme**. Onsets are optional in many languages, but rhymes are not. The rhyme can also be subdivided into two parts: the **nucleus** and **coda**. In a stereotypical syllable (something like the English word *dog* [dɑg]), the onset is a consonant ([d]), and the rhyme is the rest of the syllable ([ɑg]). Within the rhyme, the vowel is the nucleus ([ɑ]) and the last consonant is the coda ([g]). Syllables are often diagrammed like this:



The symbols above come from Greek and are as follows: σ = syllable; ω = onset; ρ = rhyme; v = nucleus; κ = coda. In building your conlang, you will have to decide what syllables are allowed. For example, all languages allow onsets, but some, like Arabic, require them. Many languages allow **closed syllables**—i.e. syllables with a coda—but not all of them do (Hawaiian requires **open syllables**, syllables without a coda). *All* languages have rules about what sounds are allowed in the onset, nucleus and coda positions.

In addition to syllabic restrictions, there are also word boundary and word-internal restrictions. In discussing these restrictions, I'll need to introduce a new bit of transcription. A period [.] is used to separate syllables. In future, all phonetic transcriptions will be broken down syllabically. Going back to boundary restrictions, in

English the sound [ŋ] isn't allowed to begin a word, but it may begin a syllable. So while there will never be a word of English that begins with [ŋ] (cf. the Nguyen example above), the word *strongest* is best syllabified [stɹɑ.ŋəst].

Since I transcribed the word, though, this is a good time to talk about **ambisyllabicity**. Certain sounds in certain languages are often sound to be neither a coda nor an onset—or both. For example, [ŋ] ordinarily isn't an onset, so in *strongest*, one almost wants to say that it's both coda *and* onset. This happens a lot with [ɹ], where in a word like *baron*, it's hard to say if it should be [bɛ.ɹən] or [bɛɹ.ən], because the [ɹ] is clearly having some sort of appreciable effect on the vowel the way a [t] or [s] wouldn't. This is generally a concept you can leave alone, but I would like to use it to introduce the concept of a **geminate**. A geminate is to a consonant what a long vowel is to a vowel. Compare the [s] sound when pronouncing *Miss Ally* and *Miss Sally*. The [s] sound should be quite a bit longer in the second. In effect, that is what a geminate is, any consonant can be a geminate. Certain languages treat geminates as the incidental co-occurrence of the same sound (as in English), but others treat geminates as longer versions of a single consonant—i.e. a consonant that is required to occupy both the coda of the previous syllable *and* the onset of a following syllable. We'll revisit this concept later on.

Many languages will place extra selectional restrictions on onsets and codas. For example, Japanese only allows [n] as a coda. In Dothraki, a largely permissive language when it comes to consonants and consonant clusters, the consonants [g], [q] and [w] can't end a word, but may end a word-internal syllable. Thus, a word like *leqse* "rat" is transcribed [leq.se]. A verb like *haqat*, "to be tired", though, which *should* be *haq*

[haq] in the past tense, turns out to be *haqe* [ha.qe], because word-final [q] isn't permitted. The addition of the [e] in Dothraki is what is known as a **repair strategy**. All language have different types of repair strategies to ensure that their phonotactic rules aren't violated. Consider all these words which came to English (ultimately) from Greek: *psychologist*, *pterodactyl*, *pneumonia*, *gnomic*. All of those words were pronounced with an initial [p] or [g] in Greek. In English, we simply can't begin a word with consonant clusters like these. Our repair strategy was to pretend those initial consonants simply didn't exist.

Other languages employ different strategies. For example, though it's fine for a word to begin with [st] or [sp] or [sn] in English, it can't do so in Spanish. No syllable in Spanish can begin with a fricative followed by a stop of any kind. Thus, when a Spanish speaker goes to a Starbucks, the way they'll pronounce it is [es̪.t̪aɾ.baks] or perhaps [es̪.t̪aɾ.boks].

The strategy employed in both situations is to preserve the canonical syllable structure of each language. Part of defining the syllable structure is defining which consonant clusters are allowed. While English is more permissive in this regard than Spanish, it's not as permissive as, say, Russian, where the Russian equivalent of *psychologist* leaves the [p] from Greek intact: психолог [psi.xɔ.lək].

In order to avoid having to pair every single consonant in one's inventory with every single other consonant, one generally uses classes of sounds (e.g. oral stops can be followed by approximants). How to decide which clusters will work and which won't, though? Let me introduce the **sonority hierarchy**. The sonority hierarchy defines

classes of sounds based on how likely they are to serve as the nucleus of a syllable. Going from least likely to be a nucleus to most, this is the sonority hierarchy:

Oral Stop > Affricate > Fricative > Nasal > Approximant > Vowel

As you can see, vowels are most likely to be the nucleus of a syllable and stops are the least likely, which should be fairly intuitive. That is, if you have a basic syllable [pa], it's pretty safe to say that, no matter what your language, [p] will be the onset and [a] will be the nucleus. It's almost unimaginable to conceive of it any other way.

Having said this, the sonority hierarchy is more like a maxim than a law. Consider that if you obey the above hierarchy precisely as written and you had to arrange three consonants—[t], [s] and [a]—into a coda-less syllable, the optimal order would be [tsa]. In English, we know that won't work, and that the non-optimal order [sta] would be preferred. Hungarian, on the other hand, is perfectly fine with [tsa], and wouldn't be okay with [sta]. This is part of what defines both languages.

That's about how you should make use of the sonority hierarchy in designing your sound system. It's not a law that tells you how sounds should be arranged, but a guideline that can be used to help you define *your* sound system. In English, for example, it's not especially worth noting that approximants like [ɹ] and [l] can follow stops and fricatives. It is worth noting, though, that [s] (and occasionally [ʃ]) can precede stops and nasals. Notice how English does that, though. Off the top of your head, you should be able to think of dozens of words that begin with [sn], [sm], [st], [sp] and [sk]. You should also be able to come up with a lot that begin with [spɹ], [stɹ], [skɹ] and [spl]. How about [skl]? There's *sclerosis*, but can you come up with any others? And how about [stl]? There are *none*. Think about how odd that is, given the classes of sounds.

What's wrong with [tl]—or [dl], for that matter? Other languages do it. One of my favorite Irathient words is [tla.nəs], which means "short visit". In English, however, it's forbidden.

By making use of and reference to the sonority hierarchy, you're able to identify what clusters and combinations are outlawed or allowed in your language that would be surprising or out of the ordinary. For example, Sanskrit allows the consonants [r] and [l] to serve as the nucleus of a syllable. Some languages are even more permissive than that. Here, for example, is a word from Georgian, with my best attempt at syllabification afterwards: გვფრცქვნი [gv.prts.kvni] "you peel us" (I'm sure there's a context where this sentence would make sense). Pretty much anything that has even the *slightest* bit of continuity can serve as the nucleus of a syllable, if you want it bad enough. Consider *pssst!* in English. It's not a word in the conventional sense, but it has clear meaning ("Hey! Pay attention to me, but don't make it look like you are!"), and is quite clearly [pṣ] with no vowel. In English, this isn't word-like enough to get word status, but why couldn't it in a different language? And, indeed, words like this can and do occur in many languages around the world.

Before leaving the sonority hierarchy, it's worth noting that if you run the sonority hierarchy in reverse…

Vowel > Approximant > Nasal > Fricative > Affricate > Oral Stop

…it gives you the end of a syllable. The thing to remember about codas, though, is that they're simultaneously *less* permissive and *more* permissive than onsets. For example, Ancient Greek allowed codas of [n] and [s] *and that's it*. Such a thing is far from uncommon. But look at English. Compared to a language like Spanish, we allow tons of

onset types, but, my stars, the things we can end a word with! Bask in some of these truly, *truly* awful codas:

strengths [stɹeŋθs]

worlds [wə˞ldz]

sixths [sɪksθs]

fifths [fɪfθs]

crafts [kʰɹæfts]

Try running any of *those* in reverse and tell me English is a simple language. The word *sixths* is especially appalling. Seriously, is that a joke? [ksθs]?! And don't get me started on English's ugliest word *crafts*.

If there's anything to take away from this, it's two linguistic tendencies. The first is that languages tend to place a lot more restrictions on codas than onsets. The second is that certain languages will pile up coda consonants, apparently because they think the word is done with and no one will notice or care.

For a conlang, though, all of this must be worked out and stated explicitly. It will help you as a language creator craft words that look and feel like they all came from the same language, and will help to prevent your native language intuitions from constraining the phonological choices you make.

Allophony

Discussing allophony is really the first step towards understanding the systematicity of language. To explain it, let me take a fun topic like werewolves and ruin it by turning it into math.

The idea behind the most common version of the werewolf is simple. Some individual (we'll call him Tony) is a regular human being, but when there's a full moon in the sky, Tony becomes a werewolf: a wolf-like man. Tony as Tony and Tony as the werewolf are never in the same place at the same time, because they're one and the same person. Furthermore, assuming that Tony acquired his wolfish second skin via some sort of bite that he received late and life, we can say that Tony is *basically* human (i.e. he started out human, most of the time he's human, and he thinks of himself as human with a bizarre ability). Furthermore, Tony doesn't become a *wolf*: he becomes a wolf*man*. So he's still basically human. If we wanted to describe Tony's two states, then, we might describe them this way:

Tony as a Human = [+human, -wolf]

Tony as a Werewolf = [+human, +wolf]

Humans are naturally [-wolf] (that we know of), so really [+human] is all you need to describe Tony. After all, we would also assume he's naturally [-bird], [-car], [-greeting card], [-oak tree], etc. If you had to describe Tony using a set of equations, then, you might do so in the following way:

1. /Tony/ > [+wolf] / _sky[+full moon]

   Translation: Tony becomes [+wolf] in the context of a sky that is [+full moon].

2. /Tony/ > [Tony] / elsewhere

Translation: Tony becomes regular Tony in all other contexts.

Here, the forward slashes make reference to a kind of meta-Tony that manifests himself in one of two ways. The greater than sign > is used as an arrow to indicate that some change occurs in Tony. In equation (1), Tony acquires the feature [+wolf]; in equation (2), there is no change, so he comes out as natural Tony. After this comes a forward slash after which is the environment that effects the change. In equation (1), the underscore _ stands for Tony, who is "occurring" (or existing) before a sky that has a full moon (in this case, it is [+full moon]). In equation (2), the environment is "elsewhere", which indicates that every other possible environment will cause Tony to remain Tony.

If you can understand the alternation between Tony and his werewolf form, you can understand allophony.

Allophony describes the *regular* distribution of sounds in a language. Some sounds rarely change, or change very little (like [g] in English), but others change quite a bit. When the change is regular and predictable, we call each instantiation of the sound an **allophone** of one **phoneme**. Using our example above, Tony, the individual consciousness, would be a phoneme, and both human Tony and werewolf Tony would be allophones of the individual Tony. To use a relatively simple example from English, let's talk about aspiration. Here's one way you could write up the aspiration rule of English:

/p, t, k/ > [p^h, t^h, k^h] / #_

/p, t, k/ > [p, t, k] / elsewhere

In other words, /p, t, k/ become aspirated at the beginning of a word (the little hashtag # mark is used for a word boundary, so #_ means when the sound in question

occurs with nothing before it), and they remain themselves otherwise. The actual details are a little more complex than this, but that's the gist of it.

Now, as an English speaker you should have the sense that there is no important difference between [p] and [pʰ]. The *p* in *peak* feels like the same sound as the *p* in *speak*; you wouldn't want to say they were different, even though they are. This is why we can refer to the phoneme as /p/ without too much trouble. Other languages experience changes just like this with their phonemes that we would perceive as rather different. For example, in Hawaiian, the phoneme /w/ surfaces as [v] when it occurs after [i] or [e]; as [w] when it occurs after [o] or [u]; and in free variation between [w] and [v] after [a]. Basically [v] and [w] are treated as realizations of the same sound. In English, the two are quite different (consider *why* and *vie*), of course, but it's the language that determines what distinctions are going to be relevant and which aren't.

In this section I'll be going over some basic types of allophonic distinctions languages make. Before moving forward, though, let me talk a little bit about how we determine when two sounds are allophones of one phoneme and when they're not. Instrumental in making this determination is the presence of **minimal pairs**. A minimal pair is like *why* and *vie* in English. These two words are pronounced identically save for the first sounds, [w] and [v]. Since the two words have very different meanings, we can determine that [w] and [v] are *not* allophones of the same phoneme. Comparing [p] to [pʰ], though, there are no words where replacing one with the other would change the meaning. This is evidence for (but not proof of) the [p] and [pʰ] being allophones of the same phoneme. It's crucial that there be no other distinctions that would complicate the analysis. For example, *pile* and *style* are different in multiple ways (presence and

absence of initial [s]; presence and absence of aspiration; different places of articulation for the stops). If minimal pairs can't be found, though, near minimal pairs can be used so long as the crucial distinction remains in tact. For example, *rep* and *let* are good enough to show that /ɹ/ and /l/ are separate phonemes, even though there's no *lep* or *ret*, because we know the endings [ɛt] and [ɛp] are distinct in English (cf. *pep* vs. *pet*), and that if a word like either *ret* or *lep* existed, it would be able to be distinguished from *let* and *rep*, respectively.

As always, though, common sense knowledge about language must be employed to determine whether it makes sense to even investigate a pair of phonemes. The famous example in English is [h] vs. [ŋ]. In Engilsh, [h] can never end a word, and [ŋ] can never begin a word. Consequently, one might draw up a set of rules like this:

/h/ > [ŋ] / _#

/h/ > [h] / elsewhere

At base, though, we know this makes no sense. There's no relationship between these two sounds in any language; English speakers do not have the sense that [h] and [ŋ] are basically the same sound; we see no evidence for word-final [ŋ] becoming [h] when a suffix is added, etc. So just because something passes the test doesn't mean it will make sense if added to a conlang. Consider this fake example:

/a/ > [ɑ] / _[i, e]

/a/ > [l] / _[u, o]

/a/ > [mp] / elsewhere

Maybe the first line would work in some language, but to have a vowel become a lateral in one instance and then a sequence of consonants in some other instance

makes absolutely no sense. The rules are well written and thorough, but it doesn't pass the smell test. There's no way something like this would work in a natural language.

Keep these facts in mind as we look at some common allophonic patterns that we see crosslinguistically. It can be difficult to distinguish between synchronic (or current) allophonic variation and diachronic (historical) sound changes, so you should use the following section in tandem with the sound change section in the Evolution chapter.

Edge Phenomena

A number of sound changes occur specifically at the beginning or ends of words. Here's a list of some of the most common changes that occur:

Word-Final Devoicing: Voiced obstruents (oral stops, affricates and fricatives) often devoice at the ends of words. This happens in languages like German, Russian and Turkish. For example, Turkish borrowed the Arabic word [ki.taːb], which means "book". In Turkish, this comes out as *kitap* [ki.tɑp]. If it becomes the object of the sentence, though, it takes on a suffix, becoming *kitabı* [ki.tɑ.bɯ], with the fully voiced [b] restored. The rule written out looks like this, where C stands for any consonant:

C > [-voice] / _#

Phoneticians say this type of variation is quite common because of the way voicing works. Since we maintain voicing by using air from our lungs to vibrate our vocal folds, it easier to voice sounds at the beginning of the word than the end, for the simple reason that we gradually run out of air as we speak. It's certainly possible to maintain voicing at the end of a word (it happens in English), but it's not uncommon for all obstruents to be voiceless at the end of a word.

Word-Initial Aspiration: We've seen this a lot already, but languages that don't distinguish between aspirated and unaspirated consonants (like English, Japanese, German and Hawaiian) may feature aspiration word-initially and not necessarily elsewhere. Aspiration is actually the delay of the onset of voicing (since vowels are usually voiced), so word-initial aspiration is actually a way of

delaying the use of the vocal folds until it's absolutely necessary, and so it's more common at the beginning of a word than elsewhere. Plenty of languages distinguish aspirated and unaspirated consonants as phonemes, though, so it's not at all necessary.

Word-Final Vowel Neutralization: However many vocalic distinctions a language makes, it's not uncommon for that distinction to be neutralized at the end of a word in one or more ways. For example, later speakers of High Valyrian neutralized the distinction between long and short vowels at the end of a word. Here's an illustrative example:

vali /va.li/ [va.li] "men (subject of a sentence)"

valī /va.liː/ [va.li] "men (object of a sentence)"

In addition, older Castithan disallowed high short vowels at the end of a sentence. Here are a couple examples:

 /ku.vi.ku me/ [ku.vi.ko me] "for a spouse"

 /ku.vi.ki me/ [ku.vi.ke me] "towards a spouse"

Some languages will do the exact *opposite* of this, only allowing "cardinal" vowels in word-final position (e.g. if a language has the vowels /a, e, i, o, u/, it will only allow [a, i, u] in word-final position). Other languages will take any vowel—or often just the non-high vowels—and reduce them to [ə] in word-final position. The reasoning for this is generally the same as with word-final devoicing: there's less air at the end of a word than the beginning, and it's often more economical to make fewer distinctions at the end of a word than at the beginning.

Word-Final Consonant Neutralization: Some consonants experience radical changes at the end of a word, but the changes are sometimes language- or even dialect-specific and a bit hard to characterize. For example, in Cuban Spanish, all word-final nasals are realized as [ŋ], whereas in many other varieties of Spanish, all word-final nasals are realized as [n]. Many languages will allow a word-final [ʔ] to replace a word-final voiceless consonant with no apparent change in meaning or perception. This can be heard in the Cockney pronunciation of a word like *lot* [lɔʔ], but is also quite common cross-linguistically—as is the spontaneous appearance of word-final [p] for similar reasons (this is the theory behind the words *yep* and *nope* in English). In all cases, though, the original sounds are simplified to a single realization—and one which usually can be explained by other typological, acoustic or phonetic principles.

Intervocalic Phenomena

Any time a consonant appears in between two vowels there's potential for **lenition**. Lenition is a general cover term that means a sound is "simplified" in some context. In the case of a vowel-consonant-vowel (VCV) sequence, vowels are two very similar things, and a consonant is a very different thing. It's not uncommon for the consonant to become more like the vowels on either side of it simply as a matter of course—and often the change isn't even noticed by the speaker.

Every instance of allophonic variation you will see below is a form of lenition. Keep the description above in mind as you look at each one:

Intervocalic Voicing: Vowels are usually voiced, so if a voiceless consonant occurs between two voiced vowels, a common change is to voice that consonant. The rule looks like this:

C > [+voice] / V_V

Intervocalic voicing is *extremely* common, and also *extremely* unstable. For example, it used to be a regular pattern of English (this is why we have pairs like *south* [sawθ] and *southern* [sʌ.ðə˞n]), but it isn't anymore, as we have no problem with pairs like *kiss* [kɪs] and *kisses* [kɪ.səz] or *bake* [bek] and *baker* [be.kə˞]. This happens with a lot of languages. What begins as a synchronic change becomes a diachronic change. It appears to be mostly synchronic right now in Italian with /s/ becoming [z] in between vowels with a few exceptions, but who knows how long that will be the case. It's common for such changes to affect smaller classes of sounds than it potentially could, as with /s/ in Italian. It's also common for the change to be sensitive to word boundaries. For example, while the change was

current, we had *knife* vs. *knives* in English, but *fire* and *afire*. Somehow languages can determine that *fire* is still the same word, so adding a prefix to it "doesn't count" as changing the phonological environment. Whether or not this will happen depends on how serious the language is about adhering to the rule, and how long its influence lasts. For example, in Castithan, the practice remained synchronic for quite a while, so it's not uncommon to find pairs like this one:

 /ʃi.no.lu/ [ʃi.no.lu] "to advance"

 /ʃi.ʃi.no.lu/ [ʃi.ʒi.no.lu] "to transcend"

Above, the word *shinolu* is modified with the prefix *shi-*, but the voicing still occurs. Languages will differ from other languages in how pervasive the rule is, and will also differ within their own vocabulary, depending on how speakers think of the word (i.e. whether the derived word is actually an independent word, or if it's really just a modification of another word that must remain intact in order to be recognizable). The same goes for other processes that would affect the recognizability of a root.

<u>Intervocalic Spirantization:</u> This change takes a stop and turns it into a fricative of some kind. The rule looks like this:

C > [+continuant] / V_V

It's quite common, crosslinguistically. One of the most frequently cited examples comes from Spanish, where all voiced stops become fricatives in between vowels. My favorite illustration of this phenomenon is the word *abogado* "lawyer" which is properly pronounced [a.βɔ.ɣa.ðɔ]. Elsewhere, though, the stops remain

stops, as in *bomba* "bomb" [bom.ba] and *grande* [gran.dɛ]. This is a true synchronic phenomenon as there are no instances where the rule does *not* apply, and there isn't even a possible word where the fricatives could be contrasted with the stops. Otherwise, it's actually quite difficult to find a synchronic example, as so many diachronic examples exist. For example, compare the Latin words on the left with their English counterparts on the right (both sets of words came from the same proto-language):

pēs [peːs]	foot [fʊt]
tertius [ter.tius]	third [θəˑd]
canis [ka.nis]	hound [hawnd]
quod [kʷod]	what [wʌt] (older pronunciation [ʍat] or [hʷat])

This change, known as Grimm's Law, turned *p into [f], *t into [θ], *k into [h] and *kʷ into [hʷ]. This is a classic example of spirantization which *must* have been synchronic at some stage in Germanic even if today it's merely a fossil.

Further Intervocalic Lenition: There's no good name for these types of changes, but they're equally common. Consider a true synchronic change in English: the alternation of [t]/[d] and [ɾ]. Depending on where *t* or *d* (or *tt* or *dd*) occurs in a word, an English speaker will *only* pronounce it one way: for /t/, either [t], [tʰ] or [ɾ] (sometimes [ʔ]), and for /d/, either [d] or [ɾ]. This can be tough for an English speaker to wrap their head around, since [ɾ] is an *r*-like sound, but that is what we pronounce in words like *madder*, *matter*, *adamant*, *fodder*, etc. What happens with this change is that the sound is sped up to such a degree that it becomes

the quickest version of that sound. This is how /t/ goes directly to [ɾ] without an intermediate step of [d]. Other languages will have an alternation with [l] or [ɭ] instead of [ɾ]. Certain other changes of this type have sounds like /ɻ/, /dʒ/ or /ʄ/ going to [j] intervocalically; /p/ or /b/ going to [v], [β], [ʋ] or [w]; /ʈ/ and /ɖ/ going to [ɽ]; and /k/ or /g/ going to [ɣ] or [h]. The difference between this and spirantization is minimal, but spirantization really only makes one change: from consonant to fricative. And while that affects classes (e.g. just voiced stops or just voiceless stops), this change affects places of articulation. This is why in English /t/ and /d/ go directly to [ɾ] (this is actually especially interesting with certain varieties of Canadian English, where a word like *rider* [ɾajː.ɾɚ˞] will be different from *writer* [ɾaj.ɾɚ˞] because the vowel will have failed to lengthen before the voiceless /t/ in *writer*).

As a side note, intervocalic lenition is far more common than any other type of intervocalic change. At its furthest, sounds can simply disappear, but that's something we'll address in the Evolution section.

Consonant-Specific Phenomena

Certain consonants effect certain types of changes in the sounds around them. This is a short list of some of those phenomena:

Nasalization: Coda nasals (and sometimes onset nasals) quite commonly cause nasalization in surrounding vowels. This happens routinely in English (try plugging your nose and focusing on the differences between the vowels in *pot* and *pawn*). The result of the nasalization is often not noted, but can be quite dramatic. In Southern Californian English, for example, a narrow transcription of *bat* and *ban* will produce strikingly different results: [bæt] vs. [bə̃ə̃n]. This is partly due to the effect of the nasalization, but also partly due to the acoustic effect of having most of the air—and sound—pass through the nose. The result is often a raising of low and mid-low vowels and the neutralization of higher vowels. Here's an example from Irathient, where low /a/ raises to [o] before nasal codas, with no perceived difference on the part of native speakers:

 /θak.tu/ [θak.tu] "birth"

 /θan.tu/ [θon.tu] "change"

Nasal Assimilation: One extremely common change is for languages with one or two nasal phonemes—usually /n/ and /m/—to have a nasal assimilation rule. This causes a nasal to change its place depending on the consonant that follows it. We actually saw a fairly good example of this in the section on nasals. Consider the alternations in the *in-* prefix in English:

implode [ɪm.pʰlod] *indent* [ɪn.dɛnt]

invite [ɪmˌvajt] *incapacitate* [ɪŋ.kʰə.pʰæ.sɪ.tʰet]

Essentially the nasal just becomes a nasal stop version of the next sound. Changes like this are common for a nasal phoneme /n/; less common for a separate /m/ phoneme; even less common for other stand-alone nasal phonemes like /ɲ/ and /ŋ/ which tend to remain intact. It is worth noting that sometimes a feature of a particular affix is that the nasal *doesn't* assimilate. Consider the following examples:

unpopular [ʌn.pʰɑ.pjə.lə˞] *undone* [ʌn.dʌn]

unfriendly [ʌn.fɻɛnd.li] *unyielding* [ʌn.jil.diŋ]

unthoughful [ʌn.θɑt.fəl] *uncouth* [ʌn.kʰuθ]

Rhotic Assimilation: *R*-like sounds have *r*-like effects on surrounding consonants and vowels. In vowels, so-called *r*-coloring can effect the perception of the previous vowel, resulting in a slightly different version of that vowel. Consider the number of vowel distinctions you can get before [d] in English versus those you can get before [ɻ]. Another common change *r* sounds can have on a previous or following coronal consonant is turning it into a retroflex sound. This is what happens to the flap in *mortar* [mɔ˞ɽə˞], where the ordinary /t/ > [ɾ] rule turns [ɾ] into [ɽ] on account of the surrounding *r*-colored vowels.

Back Consonant Coloring: Certain consonants which pull the tongue back towards the pharynx often affect the pronunciation of surrounding vowels. This phenomenon is associated with uvular, pharyngeal and epiglottal sounds (*not*

glottal sounds, where the tongue isn't involved). You see it in Arabic with its so-called "emphatic" consonants. Here are some contrasting examples (focus on the vowels after [k] vs. [q]; [h] vs. [ħ]; and [t] vs. [tˤ]):

كلب /kalb/ [kalb] "dog"

قلم /qa.lam/ [qɑ.lam] "pen"

هيش /hiʃ/ [hiːʃ] "thicket"

محيد /ma.ħiːd/ [ma.ħeid] "avoidance"

زيتون /zai.tuːn/ [ze.tuːn] "olive"

طول /tˤuːl/ [tˤəul] "length"

This is a near universal phenomenon that you should consider making reference to if you plan on using any "back" consonants in your language. In effect, the range of motion of the tongue lessens, so that vowels become lower and more back than they would be otherwise.

Glide Fortition: This is one of my favorite sound changes. In some languages, the glide phonemes /j/ and /w/ harden before like vowels. The form will change, depending on the language, but one version will have /j/ becoming [ʒ] before [i] and /w/ becoming [v] before [u]. Other possibilities for /j/ are [ɟ], [ʝ] and [dʒ], and for /w/, [ʋ] and [β].

Vowel-Specific Phenomena

Certain vowels effect certain types of changes in the consonants around them. This is a short list of some of those phenomena:

Palatalization: Palatalization is pervasive throughout the world's languages, though it too is a change that tends to be historical in nature. Nevertheless, there are plenty of examples of synchronic palatalization, as it's so common. Compare your pronunciation of /h/ in *heat* and *hot*. They should feel quite different. This is because if you wanted to do a very narrow transcription, *heat* would be transcribed [çi̯t] and *hot* would be transcribed [hɑ̯t]. Similar small changes accompany the production of non-labial stops before the high vowel [i] (compare *keep* and *cop*; *teal* and *tall*). In effect what happens is the tongue anticipates the production of a high vowel and so it gets ready for the vowel, affecting a preceding consonant. Which vowels will cause palatalization varies language to language, but they're usually the mid-high to high front vowels: [i, ɪ, y, ʏ, e, ø]. In rare cases palatalization will occur for other reasons—or will be caused by other vowels—but the majority of palatalization occurs with these vowels. It primarily affects consonants pronounced with the tongue, but palatalization of labials isn't unheard of.

Labialization: The counterpart to palatalization is labialization. Consonants that occur before the mid to high back rounded vowels [u, ʊ, o, ɔ] have a tendency to round before those vowels and only those vowels. This doesn't really happen in English, but it's quite common in African languages. Note that this is *synchronic*

labialization. Many languages have labialized consonants as independent phonemes, but allophonic labialization would have a distribution like this (nonce examples):

/kela/ [ke.la] ~ /kola/ [kʷo.la]

/kila/ [ki.la] ~ /kula/ [kʷu.la]

Labialization tends to affect velars and uvulars more than other types of consonants, but it's not unheard of to have other labialized sounds.

Vowel Quality and Length with Codas: Some languages allow all of their vowels of any length to occur in open and closed syllables (Arabic, for example). Others place restrictions on what may occur in an open and closed syllable. In some varieties of Spanish, for example, the phoneme /e/ shows up as [e] in closed syllables and [ɛ] in open syllables, and the same happens with [o] and [ɔ]. In the history of Germanic, there was a time when short vowels lengthened in open syllables only. We can see evidence of this in the many English words with "long" vowels that have a "silent *e*" on the end: *fine*, *bone*, *rune*, *bane*, *gene*, etc. The "silent *e*" didn't used to be silent, which meant that the previous vowel was in an open syllable. This caused those vowels to lengthen, and when the final *e* was lost, English was left with many words with long vowels in closed syllables. The opposite process—shortening of long vowels in closed syllables—happens in certain languages. In Hausa, for example, you see alternations like the following:

gōnā [goː.naː] "farm"

gōnāta [goː.naː.ta] "my farm"

gōnarmù [goː.naɾ.mu] "our farm"

Above, the long [aː] in the word for "farm" becomes short when it's put into a syllable with a coda. *All* examples in this section are highly language-specific. There is a tendency for long vowels in open syllables and short vowels in closed syllables, but it's not a rule. With allophonic variation like this, it's up to the conlanger to decide what sounds good for their language.

There will be more to say in the Evolution section of the book (including vowel harmony!), but this will give you enough of an idea how allophony works to be able to understand the phonological phenomena you see throughout the book.

Intonation

If the phonological structure of a language is the body, intonation is the blood. Intonation is why a language written down on paper looks stale, while a language spoken aloud is music. Intonation is what will take a language you create from being a construct to being a *language*.

Intonation is also *the* most difficult element of language to represent graphically.

Before getting into it, let me give you a couple English examples that will help to demonstrate what intonation is. Consider the word *subject* in the following sentences:

Linguistics is my favorite subject.

Please don't subject me to another boring lecture on linguistics.

Notice the difference between the two instances of *subject*? What an English speaker does is alter the intonation of their voice to produce the two different meanings. In effect, that's all intonation is: the principled modulation of the pitch of one's voice for semantic or pragmatic reasons. Sometimes, as above, it's used to distinguish meaning. Other times it's used to convey extra information not encoded in the words. For example, say my friend Kyn invites me and Jon to Harbor House to eat food late at night like we were still in our 20s, and he says the following:

I'm moving to Tallahassee.

Either me or Jon is liable to respond thus:

TALL-a-HASS-ee?!

Without resorting to using any other type of machinery, if you're an English speaker, you *should* recognize this intonation pattern. It's what I've taken to calling the WTF intonation. The way it works is the word has to end in a dramatic high to low pitch

contour. We do fun stuff with it depending on the number of syllables the word we're expressing dismay over has. Consider:

BO-ored?! (1 syllable)

AN-na?! (2 syllables)

ba-NA-na?! (3 syllables)

ME-nin-GI-tis?! (4 syllables)

bur-KI-na FA-so?! (5 syllables)

HOW i MET your MO-ther?! (6 syllables)

But notice what happens when the intonational pattern doesn't play nice with the natural stress pattern of the word!

a-DU-ult?! (2 syllables)

ME-ri-da?! (3 syllables)

co-MU-ni-ty?! (4 syllables)

Ahh, I love language… Looking back at the monosyllabic example, notice that in order to make the pattern work, we lengthen the vowel to accommodate the pitch contour. Neat, huh? Then in the ill-fitting examples above, we select the pattern the word should fit (1, 2 and 3 syllables, respectively), and the remaining syllable is kind of lumped on to the end (the beginning for the first example, and the end for the next two).

Intonation ends up being drastically important to language, but since we have absolutely no good way to transcribe it—and since many conlangs are rarely spoken—it often falls by the wayside for a lot of conlangers. In this section I'll point up a few things that can be done with intonation so you can get the most out of it.

Pragmatic Intonation

I opened this section with an example of a very specific intonational pattern of English. All languages have specific patterns like that, but here I'd like to talk about some more general tendencies.

Before getting to examples, there is one general note that you should always keep in mind when it comes to oral language: Human beings have a *finite* amount of breath. Anything that a language does that requires breath will be easier to do at the beginning of an utterance.

One spot where it's easy to spot distinctive intonational patterns in a language is in questions. There are two types of basic questions: Yes/no questions and WH-questions (there are also "I wonder" questions, but we'll leave those aside for the moment). A **yes/no question** is a question that calls for an answer of "yes" or "no". A **WH-question** is a question that typically, in English, has a word that begins with "wh" in it. These are questions that have *who*, *where*, *when*, *why*, *what*, *what kind*, *how* and/or *how much*. Their intonational patterns are typically different. Consider the following three-way example from English (in this case, this is my English, so your mileage may vary):

1. *David Bowie is a genius.*
2. *Is David Bowie a genius?*
3. *What is David Bowie?*

In sentence 1, the pitch starts out high and kind of gradually lowers as the sentence progresses. In sentence 2, the pitch raises throughout the sentence, rising to its highest point on the last syllable. In sentence 3, the pitch starts high and remains

high until the end, when it drops off sharply. (Also, the answers to 2 and 3 respectively are "yes" and "a genius".) This is fairly standard for American English, but not all languages will show the same intonational patterns. Here, for example, is how you'd translate those sentences into Irathient:

1. *Gyaba Deivid Bəwi.* [ɟa.ba de.vid bə.wi]
2. *Gyaba Deivid Bəwi?* [ɟa.ba de.vid bə.wi]
3. *Deivid Bəwi hazə?* [de.vid bə.wi na.zə]

The pitch contours are markedly different, so I've recorded myself saying them and included images of the pitch tracks below:

Gyaba Deivid Bəwi.

Gyaba Deivid Bəwi?

Deivid Bəwi hazə?

As you can see, each intonational phrase has a part where it rises and a part where it falls. For statements in Irathient, the sharp fall is the first unstressed syllable after the focus (in this case, *gyaba*, "genius"). In yes/no questions, there is a rise, then the pitch stays high and lightly falls off at the end of he phrase. In WH-questions, the pitch rises very high and stays there until the unstressed portion of the final word (in this

case, the last syllable), and then it drops *sharply*. As you can see, these are markedly different from English—and other languages have different patterns still. Even with the same language, patterns can different. For example, some varieties of British English typically denote yes/no questions with a *falling* intonation at the end of the phrase—the exact opposite of American English.

Although questions are the most obvious place where intonation has a role to play in language, it is utilized in other constructions, as well. Consider the intonational patterns of the underlined words in the following English sentences:

- *Him I like*.
- *I like Iron Maiden, Dream Theater, Sonata Arctica and the Decemberists.*
- *You're going home?!*
- *No, I'm going to the store.*
- *My sister, whom you all know as Natalie, loves mushrooms.*

Each of these constructions has a special intonation associated with them. There's also the simple contrastive intonation we can use with any word in an English sentence, as shown below:

- *I ate the apple*. (*Not the banana!*)
- *I ate the apple.* (*It was me, not my sister!*)
- *I ate the apple.* (*I didn't throw it away!*)
- *I ate the apple.* (*The one I wasn't supposed to!*)

Again, while different languages will use different intonational patterns for different purposes, they all do something. And while there may be some universal tendencies (e.g. rising intonation with yes/no questions), I'd go so far as to say the

patterns are *entirely* language-specific, and that there are no universal characteristics for intonation crosslinguistically, outside this one point: Changing the intonation of something marks it in some way. If it's usual to speak with a general fall, then marking something with a rising intonation will make it more noticeable, and vice-versa. How a language will treat the fact that an item is more noteworthy than usual is language-specific.

The IPA doesn't have any outstanding conventions for marking intonation in a phrase. It has two symbols: [↗] and [↘]. [↗] means the pitch basically goes up, and [↘] means it basically goes down. They can be combined to form rising and falling intonational patterns. For myself, this simply isn't enough to capture the nature of intonation in a language, so I find using these things to be more trouble than it's worth. When I'm working on shows like *Defiance*, I don't even describe intonation patterns in the materials that go to the actors. Instead, I do two things. First, I break down every phrase syllabically and use all caps to indicate "high" tone and all lower case to indicate "low" tone. Second, I record the line. The combination of those two things is usually good enough to get the right impression across.

Realistically, though, there ought to be a better mechanism for indicating pitch than what we have. Imagine, for example, if there were five different versions of capitalization! That way you could have lower case, mid-lower case, middle case, mid-upper case and upper case. Now *that's* a system I can get behind! Until then, do your best, because intonational flavoring is well worth it (read: crucial) when it comes to making a realistic language.

Stress

Somehow I avoided talking about stress until now. I deserve a chocolate for that.

Mmm… That was good.

Stress is a property of certain languages whereby some combination of pitch, vowel length and/or volume is used to lend acoustic prominence to a particular syllable. Stress is usually a property of words, but in some languages (like French) it's a property of phrases or clauses. Stress can be **lexical** (meaning that it's different for every word and has to be memorized) or **fixed** (meaning that stress can be predicted by a number of language-specific principles). The best way to understand what stress is to see how it works. Here are some English words stressed on the last or **ultimate** syllable, with stress marked using the IPA syllable for **primary stress** [ˈ]:

alone [əˈlon]

portmanteau [pɔɹt.mænˈtow]

understand [ʌn.dəˈstænd]

Here are some English words stressed on second-to-last or **penultimate** syllable:

sofa [ˈso.fə]

illicit [ɪˈlɪ.sət]

Mississippi [mɪ.səˈsɪ.pi]

Here are some English words stressed on the third-to-last or **antepenultimate** syllable:

mechanize [ˈmɛ.kə.najz]

infantilize [ɪnˈfæn.tə.lajz]

un-American [ʌ.nə.ˈmɛ.ɹə.kɪn]

In the last two sets you'll see also that *sofa* and *mechanize* are stressed on the first or **initial** syllable. Given the general pattern of English, though, it's more likely that the stress is assigned from the right edge, so that it makes more sense to say the stress in *mechanize* is antepenultimate, rather than initial. For example, if you add more syllables, it's easy to see the shift in stress, as with *mechanization*.

The mark [ˈ] indicates primary stress. In addition to this, there is also **secondary stress**, which is marked with [ˌ]. However it's realized in a particular language, secondary stress will be indicated in some lesser way than primary stress, but will have some sort of marking that will distinguish it from an unstressed syllable. Secondary stresses *tend* to radiate out from a syllable with primary stress, skipping every other syllable. For example, looking at *un-American* again, secondary stresses appear on the first and last syllables: [ˌʌ.nə.ˈmɛ.ɹə.ˌkɪn]. If you like the theory that [ʌ] only appears in stressed syllables, syllables with secondary stress count, which is what licenses [ʌ] in the first syllable, but not the second.

There are a couple of ways to design a good stress system depending on what kind of stress system you'd like to have. If you want to design a lexical stress system (i.e. stress placement is idiomatic and must be memorized), the only good way to do it is to evolve the system. The reason English has such a random stress system is because our words have lost a *lot* of sounds over the centuries, and we've borrowed a lot of words whose stresses we also borrow (sometimes). There were regular stress rules for English at one point, but with the way the language has evolved, it just threw up its hands and said "whatever". English pretty much just sits around all day unwindulaxing

on the couch in PJs watching reruns of *Chuck* (in other words, English is a lot like me). This is why English words are stressed all over the place: They derive from a *ton* of different regular patterns. To produce something like English, you have to emulate that history. You could also just decide randomly what words are going to have stress, but the result will be artificial and unimpressive.

Fixed stress systems are a lot of fun. Certain stress systems in the world are very simple. In Finnish, for example, stress is always on the first syllable, and the first syllable is always special in some way. If you look at Finnish, you'll notice a lot of the time the first syllable has either a long vowel, a diphthong, or a coda consonant. That's not an accident. Finnish *loves* its initial stress system—so much so that certain dialects will actually geminate a following consonant if the first syllable is light and the next syllable is heavy. If anything, in Finnish placement of secondary stress is more interesting than placement of primary stress.

Aside from Finnish and languages like it, most language have a series of complex rules to determine where primary stress is placed. For example, in Arabic, primary stress goes to the rightmost heavy syllable in the root (with the caveat that word-final case vowels have been dropped in many cases). If there are no heavy syllables, it goes to the antepenultimate syllable. Here are some examples:

سيارة /sa.jaː.ra/ [saˈjaː.ra] "car"

قلم /qa.lam/ [ˈqɑ.lam] "pen" (*lost final vowel*)

كتب /ka.ta.ba/ [ˈka.ta.ba] "he wrote"

كتاب /ki.taːb/ [kiˈtaːb] "book"

زيتون /zai.tuːn/ [zeˈtuːn] "olive"

طفولة /tˤa.fuː.la/ [tˤa.ˈfuː.la] "childhood"

Most of the time it's fairly predictable. But aside from simply saying what happens, how does this stuff actually *work*? How do Arabic speakers intuitively know what syllable to stress?

The best way I've seen to analyze fixed stress systems is with a linguistics framework called Optimality Theory (OT). Eric Baković was my OT instructor at UCSD, and though I'm not convinced that the framework is applicable to all areas of phonology, I think it works astonishingly well for fixed stress systems and level tone systems (more on those in the next section). There isn't time to do a full introduction to OT here, but if you're interested in seeing just what you can do with stress in a language, I recommend looking into it.

The main idea behind OT is that there are competing forces at work in any language, and what we produce is the *least bad* version of the language based on the various competing principles in our heads. So, with an Arabic word like /sa.jaː.ra/ that could be stressed on any one of those three syllables, the version with penultimate stress, [sa.ˈjaː.ra], is the least bad—or optimal—candidate. Why? Because Arabic likes to have the stress is close to the right edge as possible, but also likes heavy syllables to be stressed. Stressing those heavy syllables, though, is more important than getting the stress as far right as possible, though, so [sa.ˈjaː.ra] is better than the impossible *[sa.jaː.ˈra] (we use an asterisk * to indicate that a form is ungrammatical or unattested).

In order for you to create your own systems using competing candidates, you'll have to some tools at your disposal to work with. Below is a list of theoretical constructs that are particularly helpful in constructing tense systems:

- Foot: A foot is a prosodic unit that many languages find useful in evaluating stress. A standard foot is composed of two light syllables, though it may have more. For example, *sofa* is a foot in English, and we write it in parentheses: (so.fə). *Mississippi* is two feet: (mɪ.sə)(sɪ.pi).
- Trochee: A trochee is a foot that has stress on the first syllable. For example, (ˈso.fə) is a trochee in English.
- Iamb: An iamb is a foot that has stress on the ultimate syllable. For example, *ago* (ə.ˈgow) is an iamb in English.
- Dactyl: A dactyl is a foot that has three syllables and initial stress. For example, *Canada* (ˈkʰæ.nə.ɹə) could be analyzed as a dactyl in English.
- Anapest: An anapest is a foot that has three syllables and final stress. For example, *Illinois* (ɪ.lə.ˈnɔj) could be analyzed as an anapest in English.
- Amphibrach: An amphibrach is a foot that has three syllables and penultimate stress. For example, *Nevada* (nə.ˈvæ.ɹə) could be analyzed as an amphibrach in English.
- Mora: A mora is a timing unit that can be useful in building stress systems. An open syllable with a short vowel is equal to one mora. An open syllable with a long vowel is equal to two mora. A closed syllable with a short vowel is also equal to two mora. Basically, each vowel unit is one mora, and each coda consonant is one mora. Some languages don't count moras beyond two, but some do. It's up to you to decide what works for your language.

All right, I know this probably seems like a bunch of gobbledygook unless you're familiar with poetics, but here's what this buys you. Let's make up some words (it doesn't matter what they mean at present, just that they have different shapes):

ka.la	hem.bek	sa.va.lon	bi.se.lu.li
pu.lim	a.ri.la	ir.gu.des	am.pe.ro.gu
tam.ba	bon.du.le	o.wek.tu	a.ten.do.run

That's enough for now. Where are these words stressed? If it's lexical stress, it's wherever the dictionary says they are. If it's a super-regular system like Finnish, it's always the same syllable. If it isn't, this is where all our parameters come into play.

For example, let's say that in this language…

1. A foot may consist of two syllables at most.
2. Extra syllables are not footed.
3. Feet are built from the right edge to the left.
4. Feet are iambic.
5. Main stress is on the right-most foot; secondary stress is placed on others.

Here's our list again with stresses marked:

(ka.ˈla)	(hem.ˈbek)	sa.(va.ˈlon)	(bi.ˌse.)(lu.ˈli)
(pu.ˈlim)	a.(ri.ˈla)	ir.(gu.ˈdes)	(am.ˌpe.)(ro.ˈgu)
(tam.ˈba)	bon.(du.ˈle)	o.(wek.ˈtu)	(a.ˌten.)(do.ˈrun)

The result is main stress is always on the last syllable and secondary stress on the antepenultimate syllable *except* in trisyllabic words. If you had a phonological rule that was sensitive to stress (like the [ʌ]/[ə] alternation we've seen in English), trisyllabic forms would behave different from tetrasyllabic forms.

Now let's change *one* thing. Let's say instead of main stress being on the *right*-most foot it's on the *left*-most foot. Here's what happens:

(ka.ˈla)	(hem.ˈbek)	sa.(va.ˈlon)	(bi.ˈse.)(lu.ˌli)
(pu.ˈlim)	a.(ri.ˈla)	ir.(gu.ˈdes)	(am.ˈpe.)(ro.ˌgu)
(tam.ˈba)	bon.(du.ˈle)	o.(wek.ˈtu)	(a.ˈten.)(do.ˌrun)

And look at that! Everything is pretty much the same except that suddenly when you get to tetrasyllabic words, primary stress falls on the antepenultimate syllable. And all we did was make one *tiny* adjustment. What if we made some massive adjustments? Say something like…

1. A foot may consist of two moras at most. If the proposed foot would have more than two moras, a light syllable may be footed by itself.
2. Extra heavy syllables are footed; extra light syllables are not.
3. Feet are built from the left edge to the right. Don't skip any feet.
4. Feet are trochaic.
5. Main stress is on the left-most foot; secondary stress is placed on others.
6. Short vowels bearing main stress are lengthened.

Here are the results:

(ˈkaː.la)	(ˈhem.)(ˌbek)	(ˈsaː.va.)(ˌlon)	(ˈbiː.se.)(ˌlu.li)
(ˈpuː.)(ˌlim)	(ˈaː.ri.)la	(ˈir.)(ˌgu.)(ˌdes)	(ˈam.pe.)(ˌro.gu)
(ˈtam.)ba	(ˈbon.)(ˌdu.le)	(ˈoː.)(ˌwek.)tu	(ˈaː.)(ˌten.)(ˌdo.)(ˌrun)

Now *that's* a pretty crazy language. For a fun exercise, try modifying the first rule above to say that if a foot would consist of more than two mora, add an [a] to produce a better foot (so /a.ten.do.run/ would become /a.te.na.do.run/). Another fun one is if the

first foot you're able to build would have more than two moras, skip the first syllable. Or say that light syllables bearing *any* stress lengthen. Try out trisyllabic feet. Have fun with it!

As with most linguistic frameworks, OT overgenerates, but for conlanging, that turns out to be a good thing, as it gives the conlanger more options. If you work with this right, though, whatever system you come up with should be *theoretically possible*, even if it doesn't exist in the real world.

Allophonic rules that are sensitive to stress tend to work hand-in-hand with stress-placement rules, unless the effect is purely phonetic. So, for example, vowels becoming long in stressed syllables, vowels reducing or disappearing in unstressed syllables, geminating a following consonant to make a heavy syllable in a stressed syllable—all of these types of changes occur because the point of stress is to make the stressed syllable prominent in some way. Use that and all these rules to your advantage in designing a system you like the sound of.

Tone

You've probably heard a thing or two about tone languages. You've probably heard languages like Thai and Vietnamese described as "musical", "sing-song" and "exotic". It may also be the case that you've never heard of the tone languages of Africa—or America. In this section, I'll give you the basics on tone, and give you an idea where to start if you want to create a tonal language.

First, **tone** in linguistics is the uniform association of pitch with phonological material used to distinguish meaning (either semantic or grammatical). The most famous example comes from Chinese, where four words with roughly the same phonetic representation—[ma]—have four different tones, and, consequently, four different meanings:

媽 *mā* [ma˥˥] or [ma55] "mother"

麻 *má* [ma˧˥] or [ma35] "hemp"

馬 *mǎ* [ma˨˩˦] or [ma214] "horse"

罵 *mà* [ma˥˩] or [ma51] "scold"

There's also a fifth *ma* that gets its tone from context only. The four words above use the IPA characters associated with tone. The characters have a vertical line as a standard with a left-pointing bar that indicates the approximate level of the tone, with low being low and high being high. The characters are (in order of lowest to highest): [˩], [˨], [˧], [˦], [˥]. They can be put in order to indicate contours or vowel duration. These characters are often replaced with numbers, using 1 through 5 going from lowest to

highest. Personally, I prefer the number system, so I'll be using it throughout the rest of the book.

The first important thing to know about tone is that it is *not* the same as a musical scale. A note of G is a specific sound which can be further specified if you indicate the octave. Tone level in a tone language, though, is relative. If a word like [ma55] has a high tone associated with it, the tone will *always* be higher than a word with a dipping [214] tone in the same utterance, but it will *not* always be the same exact pitch. Such a thing requires perfect pitch, which not all people have. Furthermore, it's unnecessary. Maintaining relative tone levels throughout a discourse is enough to distinguish meaning, and that's what's important.

Second, regarding the number of tones, natural languages can run as high as nine, and as few as two. If there are two tones, they're always high and low. Languages with more than three tones will have a contour tone of some kind. Languages that are claimed to have more than nine tones usually can't produce minimal pairs with all examples (e.g. some tones will only occur with certain codas while others only occur without codas), so their status it debatable. It is theoretically possible for a human to distinguish each pitch that an ear can hear, but languages never come close to exploiting the physical limitations of humans. This is something that can be explored for engelangs, but such a language would prove prohibitively challenging to use and understand.

Tone languages themselves seem to come in two varieties: Contour tone languages and level or register tone languages. I'll address each type separately.

Contour Tone Languages

Though I'll be introducing them here, I will start off with this caveat. If your intent is to create a contour tone language, it *must* be evolved. Based on what we know about the evolution of contour tone languages, there's no way to do it faithfully without having a full history behind it. The same isn't necessarily true of register tone languages.

With that out of the way, a **contour tone language** is a language that typically assigns a specific tone to a specific syllable, and that tone is fixed. Contour tone languages usually have at least four tones, and sometimes as many as nine. For a contour tone language, the word "tone" itself will often apply to a tone melody, rather than an actual pitch level. Most (but not all) contour tone languages in the world can be found in Southeast Asian. We've already seen the tones of Mandarin. Here's an example from Thai:

นา [na:33] "paddy field"

หน่า [na:21] "nickname (i.e. this is someone in particular's nickname)"

หน้า [na:51] "face"

น้า [na:45] "maternal aunt"

หนา [na:14] "thick"

Contour tone languages typically name the tones that are present in the language, and it's known that a particular word has a particular tone. This is actually quite different from register tone languages, where tones may change for grammatical purposes, or when affixes are added. They're unique in their construction. It's also not

uncommon for contour tone languages to have an overabundance of monosyllabic roots and words.

Though each syllable has a fixed tone, there are two exceptions to how tone is realized. Some words (*especially* function words) have no particular tone and simply adopt a tone based on what word precedes them. In addition, the tone of other words that *do* have specific tones will change depending on what words precede or follow them. Both of these phenomena are instantiations of what we call **tone sandhi**. Tone sandhi rules differ from language to language, but they describe what happens when a word with a particular tone comes in contact with another word with a particular tone. Here's an illustrative example from Chinese:

你 *nǐ* [ni214] "you" + 好 *hǎo* [hao214] "good" = 你好 *ní hǎo* [ni35.hao21] "hello"

In Chinese linguistics, this is an example of the change that occurs when two "3" (or dipping) tones come togethers. The first "3" tone becomes a "2" (rising) tone, and the second "3" tone loses its rising intonation at the end. There's nothing that would force this to happen phonetically: it just does.

If you're interested in creating a contour tone language, it behooves you to examine the tone system of various natural languages. Essentially, though, the sandhi rules are there for two reasons. The first reason is it makes pronunciation simpler. For example, in the contour tone language I created a while back called Sheli, the name of the language derives from the word for "ocean". That word is formed in this way:

 [ʃeː42] "vastness" + [liː24] "blue-green" = [ʃeː52.liː33] "ocean"

The word is a compound of one word with a falling tone followed by another word with a rising tone. When combined, the falling tone starts out a little higher, and the

rising tone is just leveled. It's higher than the end of the falling tone, so you still hear a rise, but there's no need for it to be dramatic. This is good enough. In tone sandhi, I think the attitude is often "enough is as good as a feast".

The other reason you see sandhi phenomenon is simply to indicate (or perhaps as a result of) compounding. You see this throughout languages, tonal or not. Compare the following, for example:

black bird (as in "I saw a black bird the other day. I think it was a crow.")

blackbird (as in "I saw a grey-winged blackbird the other day.")

Notice how these two are stressed differently? The second is a simple compound with one main stress: [ˈblæk.bə˞d]. The other is an adjective-noun compound with two distinct stresses: [ˈblæk ˈbə˞d]. The compound *blackbird* gets one main stress because it's being treated as one word. The same is true of compounds or phrases in many contour tone languages. If a cluster of words is supposed to go together, their tones will mingle a little bit—hence the sandhi rules. It's really no different from what happens in non-tonal languages; it just applies to the tones since they're there.

For more on evolving a tonal system see the section on Evolution at the end. Now we'll move on to register systems.

Register Tone Languages

Register tone languages have at most four tones, and often have three or two. They are sometimes called level tone languages, because many such languages are analyzed as having *only* level tones, with all contour tones simply being combinations of the other level tones. Most register tone languages have a high (H) and low (L) level tone, though some will also have a mid tone. Those that have contour usually have only a falling (HL) tone or sometimes also a rising (LH) tone.

In many register tone languages, the low tone is treated as default, with other tones being treated as special. For example, some register languages put restrictions on how many high tones may appear in a root. Others may only permit certain tone contour patterns in a word. A common pattern is to allow only all H tones, all L tones, HL or LHL. These types of restrictions vary widely from language to language, though, so it's important to take a look at a variety of level tone languages if you're interested in producing one.

Here are some examples of words in the register tone language Hausa (tone pattern indicated in parentheses at the end and with diacritics; long vowels not marked):

shekára [ʃeː.ka.raː] "year" (LHL)

shekarú [ʃeː.ka.ruː] "years" (LLH)

surúká [su.ru.kaː] "mother-in-law" (LHH)

surukúwá [su.ru.ku.waː] "mothers-in-law" (LLHH)

lábabá [la.baː.baː] "to sneak up on" (HLH)

kwáná [kwaːnaː] "nighttime" (HH)

kwanakí [kwaː.na.kiː] "nighttimes" (LLH)

da [daː] "if" (L)

dâ [daː] "previously" (HL)

I included a couple singular and plural pairs so you can see how the tones will shift around depending on the grammatical category of a word. This is far from uncommon. Take a look at these forms:

tábbatá [tab.ba.taː] "to confirm" (HLH)

tabbatáccé [tab.ba.tat.tʃeː] "confirmed (singular adjective)" (LLHH)

tabbatattú [tab.ba.tat.tuː] "confirmed (plural adjective)" (LLLH)

tábbátár [tab.ba.tar] "to confirm (used with auxiliary)" (HHH)

The tones are moving about, but it's clear that all of these are coming from the same root. This isn't something you'd see in a contour tone language. In effect, the difference between a contour tone language and a register tone language the same as the difference between a lexical stress language and a fixed stress language. And, just as OT can be good for designing fixed stress systems, so can OT be good for designing register tone systems. To do so, you'll need to decide a few issues:

1. Inherent Tone: Will certain words have inherent tone? Which tones will be inherent: Just high? High and low?
2. Inherent Tone Melodies: Will tones or tone melodies be inherent? If the latter, which melodies? H, HL, LH, L, or more?
3. Default Tone: What's the default tone? What happens to syllables that don't have tone?
4. Repair Strategies: What happens when an affix with an inherent tone is added to a word and the result is an infelicitous melody? For example, say a

H tone suffix is added to a word with a HL melody and you don't allow HLH as a melody. What happens?

5. Contours: What happens when two tones are assigned to one syllable? What contour tones are allowed? If one is disallowed, what happens when that melody occurs on a single syllable?

How you answer these questions will determine the character of your register tone language. Also, it's important to note that tone realization can be sensitive to the tones of words. For example, the Hausa copula takes the *opposite* tone of whatever it follows:

Sárkí ne. [saɽ.kiː neː] "It is a chief."

Yáro né. [jaː.roː neː] "It is a boy."

In Njaama, one of my languages, subject and object pronouns are distinguished by a change in tone (*tekaané* means "saw" below):

Wa (L) *tekaané yáá* (H). [wa te.kaː.ne jaː] "I saw you."

Yaa (L) *tekaané wá* (H). [jaː te.kaː.ne wa] "You saw me."

It really depends on the language what the tone will or won't do. Looking at a plethora of level tone languages can be confusing, so it's ideal to take a look at one, figure out exactly how it works, and then move on to another. Everything that's been discussed in this section is a possibility, though. In my opinion, register tone languages are among the most interesting languages I've seen, and I'd love to see more register tone conlangs.

Sign Language Articulation

Advance warning: This subject requires *its own book*. I'm going to try to fit as much as I can into a couple pages.

First, a **sign language** or **manual language** is one which uses the hands as its primary articulators. Sign languages also use facial expressions, eyebrow location and other parts of the body in articulation. Sign languages have been in existence for as long as humans have had language and deafness has existed. Sign languages are full systems and are complete languages; they are *not* the same as gesturing that occurs in spoken language or "body language". Sign languages have the same expressive power as any spoken language—and, indeed, can do some incredible things that spoken languages cannot. The sign languages that exist in the world today are not based on spoken languages or are in any way subordinate to them. This means that American Deaf signers that can read English are bilingual. Sign languages have their own histories that are independent from the histories of spoken languages. For example, modern French Sign Language (FSL) and modern American Sign Language (ASL) both have a common ancestor in Old French Sign Language (OFSL). British Sign Language (BSL) is unrelated to either. Sign languages are not used exclusively by deaf individuals. Many children of Deaf adults (CODAs) are fluent signers, as are other hearing individuals that have Deaf family members. Also, in American Deaf culture, *deaf* with a lower case *d* refers to the inability to hear; *Deaf* with an upper case *D* refers to the ability to sign.

That's basically the first few weeks of a Deaf culture class in one paragraph. Now onto the languages themselves.

After I signed on to be David Perlmutter's TA in his Deaf Culture course at UCSD, I found it disappointing, but not at all surprising, that no one had created a conlang sign language (CSL). The history of Deaf signing always lies somewhere below the surface of common knowledge. The lack of languages, though, wasn't due so much in part to lack of interest, but due to a lack of ways to represent them. Video remains the best way to record and transmit a sign language when face-to-face communication is impossible, but it's not (yet) convenient. This led me to create the Sign Language IPA (SLIPA), which was an attempt to encode the phonological structure of signed languages in ASCII.

Though the analogue isn't perfect, sign languages can be described in roughly the same way as spoken languages if you assume the following:

Places (P) = Consonants

Movements (M) = Vowels

Handshape (HS) = Tone

Independent Hand Movement (IHM) = Secondary Articulations

So, for example, a word in a signed language is never *just* a handshape at some location. For example, in ASL, the word for "tree" involves placing the arm straight up and down and wiggling the open hand back and forth. The wiggling is what licenses the sign. Without that, some form of movement would be needed.

In order to describe signs, then, I came up with this framework (think of this as a syllable):

p[HS]Mp[HS]

Places are always lower case; movements are always upper case; handshapes are always upper case and contained within brackets. Then syllables are described as either starting or ending at a place via some manner of movement. A word can also be a place by itself if it has a secondary articulation.

So, to use a simple example, the word for "king" in ASL is as follows: The signer makes a [K] handshape (the ring and pinky fingers are tucked into the hand; the middle finger bends forward; the index finger is extended straight up; the thumb touches the middle finger) then touches their left shoulder (s_h). After that they pull their hand to their right hip (bl) in a low arc (E^G), as if they're describing the path of a sash. To write that sign, you'd do the following (describing it from the signer's perspective):

s_h[K]E^Gbl

It looks like something you'd see in a spam e-mail, but it does the trick. SLIPA isn't the only transcription system that exists, but it's the only one that will work with ASCII. For more, type "David Peterson SLIPA" into Google.

Now for what sign languages can do. First, they're very much like spoken languages, in that they describe the world around them using nouns and verbs and place them in a particular order. Just like spoken languages, which rely on linear order, sign languages have phenomena like affixation. For example, the words for teacher and student each involve the signs for "teach" and "learn", respectively, followed by a sign where one takes one's flat hands and takes them from the shoulders down to the belt area. Thus, "teach" is a combination of "teach" and an agentive suffix, and "student" is a combination of "learn" and an agentive suffix.

Sign languages can also make use of their medium in ways that natural languages would never do. In ASL, for example, the sign for "week" requires the signer to make a [1] shape with their hand (like an English speaker holding up a number 1) and wiping it from left to right across the palm of their left hand. That's the basic sign for "week". Since handshape is used to indicate numbers, though, you can change the shape of your hand to indicate more than one week. I'm not sure how well it would work for numbers past ten, but it's a simple thing to change the handshape and sign, in one sign, "two weeks", "three weeks", "seven weeks", etc. Furthermore, since the region behind a signer is generally associated with the past and the region in front generally associated with the future, a signer can then pull their dominant hand backwards at the end of a sign to mean "ago", and push it ahead to mean "in the future". Thus, this single sign can be used to indicate between one and at least ten weeks—*and* either in the past or the future, if the signer so chooses.

The equivalent of this in spoken language would be inflection. It would be as if you could say not only *week* and *weeks* in English, but *weekso* for "two weeks", *weekt* for "three weeks", *tweek* for "three weeks ago", *fweek* for "four weeks in the future", etc. This is something that a spoken language would *never* do because it's needless complexity. Since in ASL you *already* have to use a handshape in the production of the sign, though, why not change the handshape to add more information? Sign languages routinely take advantage of the medium of sign in ways just like this to make distinctions spoken languages never would. One of my favorite examples is the sign for "understand". To form it, you face your hand towards you, raise it to your head, and then raise your index finger (kind of like a light going on). If you're really annoyed at someone

who's explaining the same dumb thing to you for the fifty billionth time, though, you can go ahead and raise your middle finger instead as a way of saying, "Yeah. I get it, dude."

This section on sign language is roughly like summarizing *War and Peace* in a haiku. If you are *at all* interested in producing a CSL, I strongly recommend you do a little investigating to see what sign languages can do so you can decide what you want to do with yours.

Alien Sound Systems

Humans can do some pretty amazing things with their mouths, hands and bodies, but everything we can possibly do is still quite human. What if there was a being that didn't have our unique physiology? How might they communicate?

At base, for language to work (as we understand it), one creature has to produce some sort of perceivable series of tokens, and another creature has to be able to perceive and decode those tokens correctly. As humans, we're able to understand and work with our five senses, so an alien language would probably need to make use of at least one of those senses, unless you're able to think up a distinct type of sense (thought doesn't count). In order to determine what makes sense, you'll first have to come up with an alien.

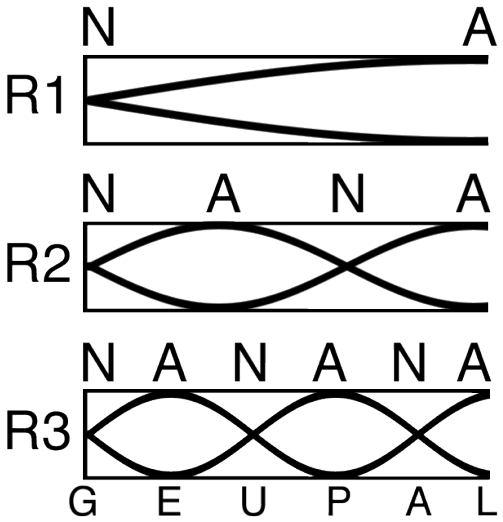
A lot of the aliens in television and film are humanoid, and differ from humans in ways that really have nothing to do with language. While Klingons have an extra set of lungs and forehead ridges, they still have one set of vocal folds, a vocal tract shaped like human vocal tracts, a tongue, an alveolar ridge, and ears. As aliens, they're simply not alien enough to warrant anything other than than a spoken human language—and the same goes for the aliens on *Alien Nation*, *Star-Crossed*, *Stargate*, *Roswell* and most of the aliens in the *Star Wars* and *Defiance* universes.

If we're focusing on speech and sound, in order to actually need different speech sounds, the aliens will need to have different vocal anatomy. How might one create a different vocal tract?

Now I'm no biologist, so I can't answer question about what is plausible, but I do know how speech sounds work, so I can give you an idea. First, recall that our anatomy

was not *designed* for language. It *allows* for it, sure, but our vocal tract is really for eating and breathing. The same will likely be true of aliens (i.e. it would be a bizarre evolutionary trend to push *specifically* for speech, since language would already have to exist. These beings would need to have language but not be able to use it, so that those that *could* use it had a distinct, evolutionary advantage. It seems implausible, at the very least). Since humans are the only creatures on Earth that can use their vocal tract to speak, then, it makes sense to take an in depth look at it.

Other animals can produce speech sounds, of course. None of them can produce all the speech sounds humans have. Part of that is due to the facility of our tongue; part of it is due to the shape of the roof of the mouth; part of it is due to what we can do with our lips. A large part, though, is due to the acoustic properties of the sound we can produce with our mouths. If the human vocal tract is modeled as a tube, this is what a standing wave's first three resonant frequencies would look like:



Key (Top and Left): N = node A = antinode R = resonant frequency

Key (Bottom): G = glottis E = epiglottis/pharynx U = uvula P = palate A = alveolar ridge L = lips

Let me explain what you're seeing here. A sound wave travels from the lungs and out towards the world. When it hits the mouth, some of it escapes, but the rest bounces off the lips and heads back towards the glottis, but with opposite polarity. This can produce a standing wave in ideal conditions. The first wave (R1) of sound has the longest wavelength, and its first overtone (R2) has half the wavelength of the original, the second (R3) has a third the wavelength, the third (R4) a quarter, etc. For speech, the first three resonances are most relevant. Where the amplitudes of the positive and negative wave are equal, that's called a node (N). Where the amplitudes are at their min and max, that's called an antinode (A). The average male vocal tract is 17 cm long.

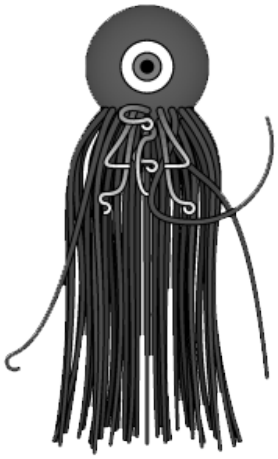
Now here's the key takeaway. If you look at where the nodes and antinodes lie for R3, they're all at key points in the vocal tract. There's a reason that a uvular stop sounds different from a velar stop, even though the surface is roughly the same. It's the combination of all these factors that allows us to produce sounds that are acoustically distinct.

If you'd like to design a different set of sounds, you'll have to design a different vocal tract—and perhaps a different set of ears. For example, while harmonics beyond the third can play a role in speech perception, it's really just the first three that are most relevant for language. If beings had better ears, they might be able to make better use of harmonics beyond the third. The vocal tract, though, is the one that will require the most work. For example, two alveolar ridges in the mouth would add an articulator, but whether or not it would sound appreciably different depends on the length of the vocal tract, and whether or not that second ridge is at a node or antinode. Different holes in the tract (like a second nose, with a passageway that led to a spot just below the

pharynx) will mean that there will be different types of nasal-like sounds available. Depending on how well the tips could be controlled, a forked tongue might not be able to produce any stop consonants. Also, imagine if a being had no tongue—or teeth—but two lips and two toothless alveolar ridges, one above and below. What type of sounds might they be able to make?

If you've got plenty of resources, you can actually *design* a working vocal draft and simulate what it would sound like if there were a being with a non-human vocal tract. This is probably beyond the realm of practicality for most people (or even productions), but it exists as a possibility.

Moving beyond sound, what if there were an alien that didn't have a mouth at all? What if it had one gigantic eyeball and forty-nine tentacles, seven of which were shorter and used as arm-like tentacles? This is the question Denis Moskowitz asked, and the language he created, Rikchik, is his answer. Rikchik is the language used by rikchiks, which look like this:



Since the seven tentacles rikchiks use to sign are the same, Denis created a language where words aren't spelled out with the signed equivalent of phonemes;

rather, words are put together that are combinations of a fixed set of shapes. Each word in Rikchik has three elements:

1. Four tentacles form the shape of a semantic category. These four tentacle are in the middle of the signing space.
2. One tentacle in the lower right hand corner makes a shape that correspond to the class of a word. This combines with the semantic shape to form a **lexeme** or word.
3. On the top of the signing space one tentacle tells you what role the word plays in the sentence (whether it's a subject or object, etc.).
4. The last tentacle, on the lower right, indicates how many of the previous signs it "collects" (i.e. how many of the previous signs go together with the current sign).

In order to be able to convey the language, Denis had to create his own transcription system, so that a word looks like this:

The four lines in the middle are the four word tentacles that indicate that the word has something to do with crystal. The small circle indicates that this word doesn't collect any others (it's just a word on its own), and the swoosh at the top indicates that this word is a quality word (meaning that it defines the quality of whatever collects it). The last symbol, the Tetris-like shape in the lower right, indicates that the type of word it is is a modifier, so basically an adjective. Literally it would be "crystal-like", but the actual definition is "sweet", since Rikchiks eat crystals as kind of a treat.

The "phonology" of Rikchik is defined by the shapes those seven tentacles can bend themselves into. While the glyph above is approximately what you will see, those are only the *ends* of the tentacles; the rest of them are connected to the body. The limitations, though, aren't breath or tongue elasticity, but tentacle elasticity (i.e. how much they can bend and into what shapes). The language that Denis has built isn't the only *possible* language that could be built using the physiology of the rikchiks; it's just the one he built.

Once you move on to things like smell, taste and touch, the question becomes what kind of language you want to build. For example, if you consider the number of speech sounds the human mouth can make, it's easy to construct languages that build words from sounds. If the number of possible sounds becomes too restrictive—or the time to produce them becomes too long—you may have to move away from words built up of arbitrary parts to a classification system like Rikchik uses. That, then, takes you beyond the realm of phonology, and into the realm of morphology and semantics.

The key point to remember is that if you're building a *truly* alien sound system (or sign system or smell system), there is no roadmap. Instead, you have to carve your own path using the basic principles we know about sound production, sound perception, object recognition, motor control, and physics. If you want to create a language for beings that can consciously change the color patterns on their wings, you need to become an expert in chromatophores. Once you know the science, then you can apply your knowledge of language to the being you've created and build on top of it.

Case Study: The Sound of Dothraki

I'm often asked what inspired me to create Dothraki, but I always find the question a little odd. The process I used for Dothraki—my first professional language—was different from any I had used before, because Dothraki was *not* my creation; the people did *not* spring from my imagination. It wasn't as if I was inspired to create a particular type of language that ended up being Dothraki. Rather, it was if I had been given a very small part of a puzzle that had been put together, and it was up to me not only to do determine what the picture was, but to create the pieces and then put it all together.

The bit that I started with was a set of Dothraki words and names from the first three books of *A Song of Ice and Fire*. This is a list of all those words (with Martin's spellings):

khal	khaleesi	khalasar	dosh	rhae
khaleen	arakh	khas	hranna	mhar
rakh	haj	rhaesh	andahli	rhaggat
dothrae	mr'anha	khalakka	vaes	dothrak
Dothraki	hrakkar	Drogo	Haggo	Cohollo
Qotho	Jhogo	Quaro	Rhaego	Ogo
Fogo	Jommo	Irri	Jhiqui	Temmo
tolorro	jaqqa	rhan	Mago	ko
Jhaqo	haesh	rakhi	Moro	Rhogoro
Aggo	Rakharo	qoy	shierak	qiya
Iggo	Zollo	Bharbo	Pono	maegi

Unless I've missed any, that's a total of fifty-five words. Ignoring grammar, the first job I had to do was figure out how to deal with the spelling. I knew going in that George R. R. Martin didn't care how fans pronounced the words and names in his book (though he did have a couple of pet peeves—like pronouncing *Jaime* [dʒem]). We didn't have George R. R. Martin as a resource when we were creating our proposals, but it also didn't seem appropriate. Since he didn't put out a pronunciation guide for his books, it was up to the fans to determine how things "ought" to be pronounced. My job, then, was to figure out how fans would most likely pronounce all the words on the list. Figuring that the bulk of fans would be English speakers—and that the executive producers, Dave and Dan were American—I decided to go with how I thought the words would be pronounced by an American English speaker.

That was the first constraint. The next was to filter that through the desire for this to be a foreign and "harsh" sounding language. That meant that, amongst other things, non-English consonants were *not* out of play. Here I was guided by the spelling. When George R. R. Martin uses spellings that are distinctly non-English, I felt that licensed the use of non-English sounds or clusters.

Finally, I was determined to treat the spelling as canon. I didn't want to change the spellings, unless it was simply to regularize them for the sake of consistency (so, for example, what was spelled *Cohollo* became *Kohollo*, since *c* is only used in the digraph *ch*). This was the rough equivalent of changing a British spelling of *colour* to an American spelling of *color*, or vice-versa. Otherwise, if two words were spelled differently, then they would be pronounced differently.

With those constraints in mind, I noticed two things:

1. The vowel *u* never occurs as a vowel; it only occurs in the cluster *qu*.
2. The consonants *p* and *b* are never used.

The only exception to the first point was in certain editions of *A Clash of Kings*, where *Vaes Tolorro* was misspelled *Vaes Tolorru*. This, though, was clearly a typo. As for the second, if you took a careful look at the table above, you'll notice that there are two key exceptions: *Pono* and *Bharbo*—the latter Drogo's father. I missed these two names entirely when crafting my proposal. This would have consequences later on.

After these realizations, I started to make some decisions about the pronunciation of the various letterforms in the extant vocabulary:

- Vowels would have their cardinal pronunciations.
- Most consonants would be pronounced just like they looked—or, at least, to an English speaker. This meant that *j* would be pronounced [dʒ] and *y* would be pronounced [j]. *Q*, however, would be pronounced [q], when occurring on its own.
- *Qu* would be reinterpreted as a sequence of [k] and [w], and would be respelled *kw* (though I also allowed the sequence [qw]).
- The spellings *th*, *sh*, *kh* and *jh* would be pronounced [θ], [ʃ], [x] and [ʒ], respectively, and the latter would be respelled *zh*. Other instances of *h* would be pronounced separately as [h], no matter where it appeared in the word.
- All vowels would be pronounced separately, even when the occurred next to another vowel. This was inspired by a phenomenon I liked in Spanish, where in a word like *creer*, there are two distinct vowel sounds: [kɾe.ˈer]. This also

meant I wouldn't have to deal with those awful *ae* sequences which are ubiquitous in Martin's works.

- *R* would work almost identically to *r* in Spanish. Phonetically, it would be a trilled [r] at the beginning or end of a word, and elsewhere it would be [ɾ]—unless it was doubled, in which case it would be [r].
- In order to maximize the distinction, all coronal consonants would be dental. If pronounced accurately, it would give the language a recognizably foreign sound.
- There would be no [u], [p] or [b]. I rationalized this by finding no instances of these sounds in the extant words (though this was a mistake!). My motivation for doing so was to give the language a unique sound (addition by subtraction, as it were). Plus, I dislike [p], [b] and [u]. I find them to be ugly sounds.

Having done this, I was able to put together the following phonetic inventory:

	Labial	**Dental**	**Alveolar**	**Palatal**	**Velar**	**Uvular**	**Glottal**
Stop		t, d			k, g	q	
Affricate				dʒ			
Fricative	f, v	θ, ð	s, z	ʃ, ʒ	x		h
Nasal	m	n	n	ɲ	ŋ	ɴ	
Glide				j	w		
Lateral		l					
Tap/Trill			ɾ, r				

	Front	**Back**

High	i	
Mid	e	o
Low		ɑ

Most of the nasal consonants would simply be allophones of the phoneme /n/ occurring in specific contexts (i.e. [ɴ] before [q], [ŋ] before [x], [k] and [g], etc.). Looking at the table I saw a couple of gaps, but the only one I decided to fill was adding [tʃ] to the palatal column to pair with [dʒ], which would be spelled *ch*. That's the only said I added to Dothraki that wasn't present in the books, based on how I interpreted the spelling of the extant vocabulary.

For the most part, I think I did a pretty good job of matching expectations for the sound of Dothraki, though there are two deviations worth noting. First is George R. R. Martin's pronunciation of the word *Dothraki*. He *consistently* pronounces it [do.ˈθɹæ.kaj]. This pronunciation is still, to me, unfathomable. I'm glad I didn't know about his pronunciation before I did my work, and I think I did right by the fans by going with what I think is the "usual" pronunciation.

One place where this didn't work, though, was with the word *khaleesi*. Obviously if you look at that word, as an English speaker, you're going to pronounce it [kʰəˈli.si]—the way it's currently pronounced by everyone—but in determining pronunciation, I had to adhere my rule that George's spellings were canon. By that rule, the word should be four syllables long, and the two *e*'s should be pronounced separately. This means its proper pronunciation is [ˈxa.le.e.si]. Now, the change from [x] to [kʰ] is to be expected (this is what we do with Greek borrowings in English, after all). If the word were *really* pronounced [ˈxa.le.e.si], though, then English speakers hearing it and turning it into an

English word would naturally pronounce it either [ˈkʰɑ.lə.si] or [kʰə.ˈle.si] (the latter as if it were spelled "kha-lacy"). Since Dothraki is a spoken language, not a written language, the pronunciation [kʰə.ˈli.si] should be impossible. It'd be like pronouncing *fiancé* [fi.ɑn.si]; it just wouldn't happen. Since we use the spelling system we do in the *real* world, though, English speakers will of course pronounce *khaleesi* [kʰə.ˈli.si]. This is the one place I wish I would have made an exception. In order to reflect the pronunciation, I should have changed the Dothraki spelling to *khalisi* and been done with it. Alas, it wasn't to be, so the gaffe will live on, to my shame.

After determining the spelling system, the goal was to produce words that looked like the extant vocabulary. For example, if you know that *band* is a word in English, and it has the structure CVCC, you should expect for there to be other words like it, and there are: *cart*, *ford*, *lamp*, *sand*, *wind*, *bolt*, etc. Part of what will give the language its character is having a bunch of words that look they obviously fit together. That's what I did with Dothraki:

Book Word(s)	**Pattern**	**Created Words**
khal, haj, dosh	CVC	*rek, qov, nith, maj, jin, has, fir, dim, chath, sash, tor*
hranna	C(C)V(C)CV	*shilla, mhotha, qwizha, rhiko, vroza, krista, hrelki*
rhaesh	C(C)VV(C)	*rhoa, noah, leik, khaor, daen, fiez, koal, mai, neak*
Aggo, jaqqa	(C)VCːV	*lorra, ricchi, ville, yalli, zajja, zhille, naffa, khirra, gillo*
tolorro, Cohollo	CVCVCCV	*zhokakkwa, najahhey, movekkha, Kovarro, inavva*

The next step was to ensure that a lot of high frequency words would have a kind of "harsh" or "foreign" sound. The first step to ensure this was to design the stress system to ensure that the rhythm would differ from that of English. In English, for

example, it's rare for a word to be stressed on the last syllable, unless the word is borrowed, or a verb. In Dothraki, all words that end in a consonant are stressed on the last syllable. Referring to the framework of the stress system, a Dothraki word looks at the right edge to determine stress. If the word ends in a consonant, the last syllable is stressed. If it ends in a vowel, it looks to the penultimate syllable. If that syllable is heavy, it's stressed. Otherwise the first syllable is stressed. This would mean that several key words wouldn't be stressed right (for example, *Dothraki* and *khaleesi* should both be stressed on the first syllable), but since they would be borrowings when spoken in English, that wasn't a big deal. What was more important was ensuring that the dialogue had that characteristic Dothraki rhythm.

So given a normal Dothraki sentence like this…

Lajak oga haz oqet ha khalaan.

[la.ˈdʒak ˈo.ga haz o.ˈqɛt ha xa.la.ˈan]

"The warrior is slaughtering that sheep for the khal."

…the stresses of the first, fourth and final words would be in the opposite place one would expect them to be if these were native English words.

The final step in making sure that the Dothraki-ness of Dothraki came through was to make sure some of the "harsh" sounds were used in high frequency terms. Remember how I talked about brand identity? This is where it becomes important. What are the sounds that were characteristic of Dothraki? In my mind, it's the doubled vowels, the voiceless velar fricative [x], geminate consonants, the trilled [r], the uvular [q] and the forceful [h], which often comes out as [ħ]. Since the audience of *Game of Thrones*

would be hearing the language through the lines, I had some control over what Dothraki words they heard. I used that to my advantage.

First, I didn't have to work very hard with [x]. Spelled *kh*, that sound is used in the words *khal*, *khaleesi* and *arakh*, which are used frequently. Just to make sure I got as much mileage out of that sound as I could, though, I made the form of one of my objective derivational suffixes -*(i)kh*. This means that words ending in [x] are *extremely* common in Dothraki: *achrakh* "smell", *mechikh* "roast quail", *nesikh* "knowledge", *sewafikh* "wine", etc.

The sound [r] was already going to be common enough, but I created two productive suffixes that end in *r*. One was derived from George R. R. Martin's word *khalasar*. This became a collective suffix that is used somewhat frequently: *astosor* "story", *gimisir* "commoners", *jereser* "market", *lajasar* "army", etc. I also created a new affix ending in *r* that turns any verb into a noun, and *that* strategy is simply ubiquitous in Dothraki. It's the strategy that turns *davra* "good" into *athdavrazar* "excellent", and *jahak* "braid" into *athjahakar* "pride".

I used a similar strategy with geminates, where turning a verb from a regular verb into a causative verb involves geminating the first consonant (e.g. *layafat* "to be happy" becomes *allayafat* "to please") and with double vowels (two grammatical suffixes attached to nouns are -*aan* and -*oon*), but [q] and [h], I did something different. Neither sound really lends itself well to morphology, so instead I simply targeted words that were either high frequency, or I knew were going to be used in a script and made sure to use [h] and [q]. For example, one of my favorite words, *mahrazh*, "man", and the word for "horse", *hrazef*, make prominent use of [h] in a position where English wouldn't

use it. A high frequency word that I knew would be used periodically is the word *qora* which is the Dothraki word for both "hand" and "arm". I also lucked out with the phrase "blood of my blood", since one of the defined words from George R. R. Martin's list was *qoy*, "blood".

After all of this, though, it was all in the hands of the actors. Those that really put a lot of effort into it and took it seriously were the ones that were really able to sell it as a living, breathing language. Jason Momoa was a gift. Not even in my imagination could I have created a better Drogo or a better flagship for this language. Up to that point, he was the hulkiest, beefiest, dreamiest mountain of a human being ever to speak a created language, and he was speaking *my* language. Try calling *him* a nerd and see how far you get! Both he and Amrita Acharia's performances sit near and dear to my heart.

Even though the process for Dothraki was a little different, since I wasn't creating a language from scratch, the principles I employed can be profitably reproduced in an original conlang. The phonetic inventory is just the starting point. The phonotactics, the stress system and the common word endings are going to be what are most noticeable to the listener, since that's what they'll be hearing the most. Use that to your advantage. By controlling those aspects of the language, you'll be defining its character, and that's what will give it a sound that is unmistakable and all its own.